

WATERMAN

(TK)

Release 3.1A John F. Collins, BioComputing Research Unit,  
Copyright (C) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPearch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run On: Mon Jun 19 16:01:00 2000: MasPar time 3.40 seconds  
Linear output not generated. 215.791 Million cell updates/sec

Title: >US-09-142-524A-4  
Description: (1-31); from USC9142524A.pep  
Perfect Score: 225  
Sequence: 1 IFSKNLNKLNMPLYIAGNKRRIKRVSNVI 31

Scoring table: FAM 150  
Gap 11

Searched: 188963 seqs, 23696106 residues

Post-processing: Minimum Match: 0%  
Listing first 45 summaries

Database: a-geneseq35  
l-geneseqf

Statistics: Mean 23.433; Variance 89.397; scale 0.252

Prof. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed.  
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID    | Description             | Prof. No. |
|------------|-------|-------------|--------|----------|-------------------------|-----------|
| 1          | 225   | 100.0       | 31     | 1 W27372 | Multi-epitope peptide   | 4.18e-14  |
| 2          | 143   | 63.6        | 20     | 1 W42130 | T-cell epitope peptide  | 9.27e-06  |
| 3          | 143   | 63.6        | 354    | 1 W04344 | Chamaecyparis obtusa p  | 9.27e-06  |
| 4          | 143   | 63.6        | 354    | 1 W42121 | Japanese cypress pollen | 9.27e-06  |
| 5          | 143   | 63.6        | 375    | 1 W04345 | Chamaecyparis obtusa p  | 9.27e-06  |
| 6          | 123   | 54.7        | 367    | 1 R45577 | Jun s I.                | 8.28e-04  |
| 7          | 123   | 54.7        | 370    | 1 R45578 | Jun v I.                | 8.28e-04  |
| 8          | 116   | 52.4        | 134    | 1 W27371 | Multi-epitope peptide   | 2.50e-03  |
| 9          | 116   | 52.4        | 353    | 1 R75366 | Japanese cedar pollen   | 2.50e-03  |
| 10         | 116   | 52.4        | 353    | 1 R41387 | Cedar pollen allergen   | 2.50e-03  |
| 11         | 116   | 52.4        | 374    | 1 R31937 | Cry j I.                | 2.50e-03  |
| 12         | 116   | 52.4        | 374    | 1 R60566 | Japanese cedar pollen   | 2.50e-03  |
| 13         | 116   | 52.4        | 374    | 1 R82490 | Cry j I Japanese Cedar  | 2.50e-03  |
| 14         | 116   | 52.4        | 374    | 1 R45541 | Cry j I pollen allergen | 2.50e-03  |
| 15         | 113   | 50.2        | 20     | 1 R45549 | Cry j I pollen allergen | 2.50e-03  |
| 16         | 113   | 50.2        | 20     | 1 R82498 | Cry j I Japanese cedar  | 2.50e-03  |
| 17         | 113   | 50.2        | 20     | 1 W44862 | T-cell epitope peptide  | 7.48e-03  |
| 18         | 87    | 38.7        | 105    | 1 W27373 | Multi-epitope peptide   | 1.89e-00  |
| 19         | 92    | 36.4        | 31     | 1 W27372 | Multi-epitope peptide   | 5.23e-00  |
| 20         | 79    | 35.6        | 80     | 1 W27369 | Multi-epitope peptide   | 7.84e-00  |
| 21         | 79    | 35.1        | 1018   | 1 W06485 | Rat contactin ligand f  | 9.59e-00  |
| 22         | 74    | 32.9        | 20     | 1 W42131 | T-cell epitope peptide  | 2.59e-01  |
| 23         | 74    | 32.9        | 637    | 1 W52294 | Drosophila p735aK prot  | 2.59e-01  |

|    |    |      |      |          |                        |          |
|----|----|------|------|----------|------------------------|----------|
| 24 | 73 | 32.4 | 238  | 1 W4755  | Human thoracic aorta G | 3.15e-01 |
| 25 | 73 | 32.4 | 298  | 1 W4727  | Human thoracic aorta G | 3.15e-01 |
| 26 | 72 | 32.0 | 963  | 1 R41082 | Thermolabile enzyme (m | 3.87e-01 |
| 27 | 71 | 31.6 | 615  | 1 R74530 | Tomato lipoxygenase    | 4.66e-01 |
| 28 | 70 | 31.1 | 613  | 1 R74532 | QTR ethylene response  | 5.65e-01 |
| 29 | 70 | 31.1 | 613  | 1 W73122 | A. thaliana ethylene r | 5.65e-01 |
| 30 | 70 | 31.1 | 721  | 1 W34454 | Racilius subtilis telc | 5.65e-01 |
| 31 | 70 | 31.1 | 746  | 1 W34455 | Racilius subtilis telc | 5.65e-01 |
| 32 | 69 | 30.7 | 20   | 1 W42129 | T-cell epitope peptide | 6.85e-01 |
| 33 | 69 | 30.7 | 635  | 1 W47124 | Tomato ethylene respon | 6.85e-01 |
| 34 | 69 | 30.7 | 635  | 1 W47125 | Tomato ethylene respon | 6.85e-01 |
| 35 | 69 | 30.7 | 635  | 1 R4633  | Nr (sever ripe) tomato | 6.85e-01 |
| 36 | 69 | 30.7 | 635  | 1 R4629  | Tomato ethylene respon | 6.85e-01 |
| 37 | 69 | 30.7 | 1029 | 1 R4627  | Homio sapiens D18511 c | 8.85e-01 |
| 38 | 68 | 30.2 | 509  | 1 R893   | Human SR-B1 class B sc | 8.85e-01 |
| 39 | 68 | 30.2 | 509  | 1 R900   | Human SR-B1 class B sc | 8.85e-01 |
| 40 | 68 | 30.2 | 1018 | 1 R4759  | Human contactin (EXHL  | 8.85e-01 |
| 41 | 68 | 30.2 | 1018 | 1 R4758  | Human contactin        | 8.85e-01 |
| 42 | 67 | 29.8 | 738  | 1 R69851 | Ethylene response (ETR | 1.01e-02 |
| 43 | 67 | 29.8 | 738  | 1 R69852 | Ethylene response (ETR | 1.01e-02 |
| 44 | 67 | 29.8 | 738  | 1 R69849 | Ethylene response (ETR | 1.01e-02 |
| 45 | 67 | 29.8 | 738  | 1 W73119 | A. thaliana ethylene r | 1.01e-02 |

ALIGNMENTS

RESULT 1  
ID W27372 standard; peptide: 31 AA.  
AC W27372:  
DT 24-MAR-1998 (first entry)  
DE Multi-epitope peptide used as immunotherapeutic agent #4.  
KW Multi-epitope peptide; immunotherapeutic agent; allergic disease;  
KW T-cell epitope region; allergen; lymphocyte; immunoglobulin F  
OS Synthetic.  
FN W09732600-A1.  
PD 12-SEP-1997.  
PR 10-MAR-1997: J00740.  
PR 10-MAR-1996: JP-C80702.  
PA (MEIP) MEIJI MILK PROD CO Ltd.  
PI Dairiki K, Iwama A, Kino K, Nume A, Sone T;  
DR WP: 97-470495/43.  
PT Peptide immunotherapeutic agent to treat allergic diseases.  
PI Contains multi-epitope pep de containing T cell epitope regions  
PI from different allergens  
PS Claim 9: Page 32: 58pp; Japanese.  
CC The present sequence represents a multi-epitope peptide which is used as  
CC a new immunotherapeutic agent. It comprises T cell epitope regions from 2  
CC or more different allergens (preferably linked via arginine or lysine  
CC dimers), where the T cell epitope regions have a positivity index  
CC greater than 100 as measured in a patient group responding to the  
CC allergen; have at least 70% reactivity with lymphocytes from patients  
CC responding to the allergen; and are not reactive with immunoglobulin E  
CC (IgE) antibodies from patients responsive to the allergen. The agent can  
CC be used to prevent and treat a wide variety of allergic diseases, e.g. by  
CC desensitisation. Side effects, e.g. those mediated by IgE, are reduced.  
SQ Sequence 31 AA:

Query Match 100.0%; Score 225; DB 1; Length 31;  
Best Local Similarity 100.0%; Pred. No. 4.18e-14;  
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Db 1 IFSKNLNKLNMPLYIAGNRRFKRVSNVI 31  
Oy 1 IFSKNLNKLNMPYIAGNRRFKRVSNVI 31  
RESULT 2  
ID W42130 standard; peptide: 20 AA.  
AC W42130:  
DT 16-JUN-1998 (first entry)  
DE T-cell epitope peptide 10 from Japanese cypress pollen antigen; Chao1.  
DE T-cell epitope pollen; antigen; T-cell epitope; Chao1; Chao2;  
KW Japanese cypress pollen; allergen; spring tree pollen disease; pollinosis.

Chamaecyparis obtusa.  
 WO9747648-A1.  
 18-DEC-1997.  
 12-JUN-1997: JP-153527.  
 (MEIP) MEIJ MILK PROD CO LTD.  
 Dairiki K, Kino K.  
 WPI: 98-052242/05.  
 T-cell epitope peptide portion of Japanese cypress pollen antigens  
 Chao1 and Chao2 used for diagnosis and treatment of spring tree  
 pollen disease.  
 Claim 1: Page 22, 71pp; Japanese.  
 The present sequence represents a T-cell epitope peptide from Japanese  
 cypress pollen antigen Chao1. The present invention describes peptides  
 which correspond to the T-cell epitope sites on Japanese cypress pollen  
 antigens Chao1 and Chao2. The peptides can be used as a reagent for the  
 diagnosis of allergy to Japanese cypress pollen, and as an antigen in  
 the treatment and prevention of spring tree pollen disease in which the  
 pollinosis involves reactivity to Japanese cypress pollen.  
 Sequence 20 AA:

Query Match 63.68: Score 143: DB 1: Length 354:

Best Local Similarity 100.0%: Pred. No. 9.27e-06: Mismatches 0: Indels 0: Gaps 0:

Matches 20: Conservative 0: Indels 0: Gaps 0:

1 IFSKNLNKLNMPYIAGNK 20  
 (|||||) (|||||) (|||||)  
 1 IFSKNLNKLNMPYIAGNK 20

## RESULT 3

ID W04344 standard: Protein: 354 AA.

AC W04344:

28-NOV-1996 (first entry)

DE Chamaecyparis obtusa pollen allergen Cha o 1 (A).

KW Pollen allergen: Cha O 1: T-cell epitope: prevention: treatment:

OS Chamaecyparis obtusa.

PN J0817510-A.

PD 09-JUL-1996.

PR 21-DEC-1994: 335089.

PA (MEIP) MEIJ MILK PROD CO LTD.

DR WPI: 96-368242/05.

DR N-PSDB: 335519.

PT DNA encoding Chamaecyparis obtusa pollen allergen Cha o 1.

PT epitope(s) of which are useful in development of preventative and

treating agent for C. obtusa pollen pollinosis

Claim 2: Pages 10-12, 17pp; Japanese.

PS The present sequence is the C. obtusa pollen allergen Cha o 1.

CC The T-cell epitopes of which can be used in the development of a

preventive and treating agent for C. obtusa pollen pollinosis.

CC C. obtusa pollen (2.4 kg) was degassed with diethyl ether, and

dried at room temp. overnight. Cha o 1 was sepd. from it and

purified. RNA was extracted from C. obtusa pollen, and mRNA and

CC cDNA derived.

CC Sequence 354 AA:

Query Match 63.68: Score 143: DB 1: Length 354:

Best Local Similarity 100.0%: Pred. No. 9.27e-06: Mismatches 0: Indels 0: Gaps 0:

Matches 20: Conservative 0: Indels 0: Gaps 0:

DB 71 IFSKNLNKLNMPYIAGNK 90  
 (|||||) (|||||) (|||||)  
 1 IFSKNLNKLNMPYIAGNK 20

## RESULT 4

ID W42121 standard: protein: 354 AA.

AC W42121:

16-JUN-1998 (first entry)

DE Japanese cypress pollen antigen Chao1.

DE Japanese cypress pollen: antigen: T-cell epitope: Chao1: Chao2:

KW diagnosis: allergy: spring tree pollen disease: pollinosis  
 OS Chamaecyparis obtusa.  
 PN WO9747648-A1.  
 PD 18-DEC-1997.  
 PR 12-JUN-1997: J02031.  
 PR 14-JUN-1996: JP-153527.  
 PA (MEIP) MEIJ MILK PROD CO LTD.  
 Dairiki K, Kino K.  
 WPI: 98-052242/05.  
 PT T-cell epitope peptide portion of Japanese cypress pollen antigens  
 Chao1 and Chao2 used for diagnosis and treatment of spring tree  
 pollen disease.  
 PS Example 1: Page 14-15, 71pp; Japanese.  
 CC The present sequence represents Japanese cypress pollen antigen Chao1.  
 CC The present invention describes peptides which correspond to the T-cell  
 CC epitope sites on Japanese cypress pollen antigens Chao1 and Chao2. The  
 CC peptides can be used as a reagent for the diagnosis of allergy to  
 CC Japanese cypress pollen, and as an antigen in the treatment and  
 CC prevention of spring tree pollen disease in which the pollinosis  
 CC involves reactivity to Japanese cypress pollen.  
 CC Sequence 354 AA:

Query Match 63.68: Score 143: DB 1: Length 354:

Best Local Similarity 100.0%: Pred. No. 9.27e-06: Mismatches 0: Indels 0: Gaps 0:

Matches 20: Conservative 0: Indels 0: Gaps 0:

DB 71 IFSKNLNKLNMPYIAGNK 90  
 (|||||) (|||||) (|||||)  
 1 IFSKNLNKLNMPYIAGNK 20

## RESULT 5

ID W04345 standard: Protein: 375 AA.

AC W04345:

28-NOV-1996 (first entry)

DE Chamaecyparis obtusa pollen allergen Cha o 1 cDNA (B).

KW Pollen allergen: Cha O 1: T-cell epitope: prevention: treatment:

OS Chamaecyparis obtusa.

PN J0817510-A.

PD 09-JUL-1996.

PR 21-DEC-1994: 335089.

PA (MEIP) MEIJ MILK PROD CO LTD.

DR WPI: 96-368242/05.

DR N-PSDB: 335519.

PT DNA encoding Chamaecyparis obtusa pollen allergen Cha o 1.

PT epitope(s) of which are useful in development of preventative and

treating agent for C. obtusa pollen pollinosis

Claim 8: Pages 11-12, 17pp; Japanese.

PS The present sequence is the C. obtusa pollen allergen Cha o 1.

CC The T-cell epitopes of which can be used in the development of a

preventive and treating agent for C. obtusa pollen pollinosis.

CC C. obtusa pollen (2.4 kg) was degassed with diethyl ether, and

dried at room temp. overnight. Cha o 1 was sepd. from it and

purified. RNA was extracted from C. obtusa pollen, and mRNA and

CC cDNA derived.

CC Sequence 375 AA:

Query Match 63.68: Score 143: DB 1: Length 375:

Best Local Similarity 100.0%: Pred. No. 9.27e-06: Mismatches 0: Indels 0: Gaps 0:

Matches 20: Conservative 0: Indels 0: Gaps 0:

DB 92 IFSKNLNKLNMPYIAGNK 111  
 (|||||) (|||||) (|||||)  
 1 IFSKNLNKLNMPYIAGNK 20

## RESULT 6

ID R45577 standard: Protein: 367 AA.

AC R45577:

13-JUL-1994 (first entry)

DE Japanese cypress pollen: antigen: T-cell epitope: Chao1: Chao2:

```

Query: Watch 52.48; score 118; DB 1; Length 124;
Best Local Similarity 60.0%; Pident 2.50e-03;
Matches 18; Conservative 5; Mismatches 6; Indels 1; Gap
DQ 78 IFSGNNIKLKPMPYIAGYKT-IDGHRREV 106
11111111111111111111111111111111
QY 1 IFSKNLNKLNPYIAGNKRPFIKRVSNW 30
11111111111111111111111111111111
RESID 9
10 P55492 standard; protein; 253 AA.
AC P55492
DE 12-WAY-1995 (first entry)
DE Japanese cedar pollen allergen (Crj) I.
DE Japanese cedar pollen allergen (Crj) I.
KK Japanese cedar pollen allergen (Crj) I.
KW prevention; treatment; cryptomeria pollinosis.
OS Cryptomeria japonica.
FH Key Location/Qualifiers
FT peptide 61..75
FT peptide 91..105 /note="T-cell epitope peptide"
FT peptide 106..120 /note="T-cell epitope peptide"
FT peptide 146..160 /note="T-cell epitope peptide"
FT peptide 211..225 /note="T-cell epitope peptide"
FT peptide 326..340 /note="T-cell epitope peptide"
FT peptide 335..346 /note="T-cell epitope peptide"
FT peptide 367..381 /note="T-cell epitope peptide"
PN JC7118295-A.
PD 09-MAY-1995.
PE 20-OCT-1993; 263626.
PR 20-OCT-1993; JP-263626.
PA (MEIP) MEIJI MILK PROD CO LTD.
PA WPI: 95-201834/27

```





OS Cryptomeria japonica.  
 WP527786-A.  
 19-OCT-1995.  
 PP 06-APR-1995: J04249.  
 PP 08-APR-1994: US-226248.  
 PK 06-DEC-1994: US-350725.  
 PK (IMMUNO) IMMUNOLOGIC PHARM CORP.  
 PA Chao X., Evans S., Franzen BM., Kuo M., Powers SP.  
 PI Shaked Z.  
 PI WPI: 95-366391/47.  
 DR N-PS33: T0248.  
 DR  
 PP: Treating Cryptomeria japonica (Cry j.) peptide(s) - useful for  
 PP: treating allergy to Japanese cedar pollen allergen or  
 PP: immunologically cross reactive allergens  
 PP Disclosure: Figure 1: 6upp; English.  
 CC Nov: peptides of cry j. have been modified as a part of a  
 CC preformulation scheme to develop an optimised drug product for  
 CC therapeutic treatment of humans suffering from allergy to Japanese  
 CC cedar pollen allergen or an allergen which is immunologically cross  
 CC reactive with Japanese cedar pollen allergen. Such modified peptides  
 CC possess certain characteristics which render them particularly  
 CC suitable for drug product formulation. Peptide fragments of Cry j.  
 CC modified and unmodified, are given in R82491-R82525.  
 SC Sequence 174 AA:  
 Query Match 52.4%; Score 118; DB 1: Length 374;  
 Best Local Similarity 50.0%; Pred. No. 2,50e-03;  
 Matches 18; Conservative 4; Mismatches 7; Indels 1; Gaps 1;  
 PI 52: HSGNNK:KLKPMYIAGYKT-FDGRGAGV 120

Q7 1 :TSKNLN:ALNMP:TAGKRF:KRVSNV 30

RESULTS

ID R45541 standard; Protein: 374 AA.

AC B45541

DI 13-JUL-1994 (first entry)

DE Cit 1 I pollen allergen

KW Japanese cedar; detection; allergy; treatment; diagnosis;

TX Cit 1 epitope; sensitivity.

OS Citriciteria japonica

Search completed: Mon Jun 19 16:01:15 2000  
Job time : 15 secs.

|    |                               |                                |
|----|-------------------------------|--------------------------------|
| KW | Japanese cedar: detection;    | allergy; treatment; glugcosis; |
| KW | t. cell epitope: sensitivity. |                                |
| KW | Civicothemia japonica.        |                                |

DE 92 IFSCNMNIKLXPMYIAGYKT-EDGRGAQV :20

\*\*\*\*\*

WATERMAN

(TM)

\*\*\*\*\*

Release 3.1A John F. Collins, BioComputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MParch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Tue Jun 20 13:34:00 2000: Maspar time 2.66 Seconds  
Tabular output not generated. 168,026 Million cell updates/sec

Title: US-09-142-524A-4  
Description: (1-31) from US09:42524A.pep  
Perfect Score: 225  
Sequence: 1 IFSKNLNKLNMPYIAGNKRRIKRVSNV1 31

Scoring table: PAM 150  
Gap 11

Searched: 143441 seqs, 14437480 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: J-issued  
1:5A\_COMB 2:5B\_COMB 3:6\_COMB 4:PCI\_COMB 5:backfiles1  
Statistics: Mean 21.952; Variance 87.801; scale 0.250

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description                       | Pred. No. |
|------------|-------|-------------|--------|-------|-----------------------------------|-----------|
| 1          | 79    | 35.1        | 1018   | 1     | US-08-452- Sequence 2, Applicatio | 6.12e-00  |
| 2          | 73    | 32.4        | 298    | 1     | US-08-118- Sequence 76, Applicati | 1.97e-01  |
| 3          | 73    | 32.4        | 298    | 4     | PCI-US93-0 Sequence 76, Applicati | 1.97e-01  |
| 4          | 72    | 32.0        | 963    | 1     | US-08-537- Sequence 3, Applicatio | 2.38e-01  |
| 5          | 70    | 31.1        | 613    | 2     | US-08-484- Sequence 48, Applicati | 3.49e-01  |
| 6          | 70    | 31.1        | 613    | 2     | US-08-484- Sequence 48, Applicati | 3.49e-01  |
| 7          | 69    | 30.7        | 525    | 2     | US-08-749- Sequence 8, Applicatio | 4.22e-01  |
| 8          | 69    | 30.7        | 525    | 2     | US-08-749- Sequence 8, Applicatio | 4.22e-01  |
| 9          | 69    | 30.7        | 635    | 2     | US-08-484- Sequence 36, Applicati | 4.22e-01  |
| 10         | 69    | 30.7        | 635    | 2     | US-08-484- Sequence 50, Applicati | 4.22e-01  |
| 11         | 68    | 30.2        | 509    | 2     | US-08-890- Sequence 2, Applicatio | 5.09e-01  |
| 12         | 68    | 30.2        | 605    | 2     | US-08-752- Sequence 8, Applicatio | 5.09e-01  |
| 13         | 68    | 30.2        | 1018   | 3     | US-08-040- Sequence 6, Applicatio | 5.09e-01  |
| 14         | 68    | 30.2        | 1018   | 1     | US-08-714- Sequence 5, Applicatio | 5.09e-01  |
| 15         | 68    | 30.2        | 1018   | 1     | US-08-408- Sequence 6, Applicatio | 5.09e-01  |
| 16         | 68    | 30.2        | 1018   | 1     | US-08-408- Sequence 6, Applicatio | 5.09e-01  |
| 17         | 67    | 29.8        | 155    | 1     | US-08-530- Sequence 12, Applicati | 6.15e-01  |
| 18         | 67    | 29.8        | 155    | 2     | US-08-484- Sequence 12, Applicati | 6.15e-01  |
| 19         | 67    | 29.8        | 738    | 2     | US-08-484- Sequence 9, Applicatio | 6.15e-01  |
| 20         | 67    | 29.8        | 738    | 2     | US-08-530- Sequence 9, Applicatio | 6.15e-01  |
| 21         | 67    | 29.8        | 738    | 1     | US-08-530- Sequence 3, Applicatio | 6.15e-01  |
| 22         | 67    | 29.8        | 738    | 2     | US-08-484- Sequence 3, Applicatio | 6.15e-01  |
| 23         | 67    | 29.8        | 738    | 1     | US-08-530- Sequence 7, Applicatio | 6.15e-01  |

|    |    |      |     |   |                                    |          |
|----|----|------|-----|---|------------------------------------|----------|
| 24 | 67 | 29.8 | 738 | 1 | US-08-530- Sequence 5, Applicatio  | 6.15e-01 |
| 25 | 67 | 29.8 | 738 | 1 | US-08-530- Sequence 3, Applicatio  | 6.15e-01 |
| 26 | 67 | 29.8 | 738 | 2 | US-08-484- Sequence 11, Applicati  | 6.15e-01 |
| 27 | 67 | 29.8 | 738 | 2 | US-08-484- Sequence 5, Applicatio  | 6.15e-01 |
| 28 | 67 | 29.8 | 738 | 1 | US-08-530- Sequence 11, Applicati  | 6.15e-01 |
| 29 | 66 | 29.3 | 152 | 2 | US-08-247- Sequence 16, Applicatio | 7.41e-01 |
| 30 | 66 | 29.3 | 152 | 1 | US-08-318- Sequence 6, Applicatio  | 7.41e-01 |
| 31 | 66 | 29.3 | 152 | 1 | US-08-795- Sequence 8, Applicatio  | 7.41e-01 |
| 32 | 66 | 29.3 | 152 | 1 | US-08-318- Sequence 8, Applicatio  | 7.41e-01 |
| 33 | 66 | 29.3 | 152 | 1 | US-08-318- Sequence 7, Applicatio  | 7.41e-01 |
| 34 | 66 | 29.3 | 152 | 2 | US-08-795- Sequence 7, Applicatio  | 7.41e-01 |
| 35 | 66 | 29.3 | 152 | 2 | US-08-795- Sequence 6, Applicatio  | 7.41e-01 |
| 36 | 66 | 29.3 | 255 | 1 | US-08-459- Sequence 4, Applicatio  | 7.41e-01 |
| 37 | 66 | 29.3 | 255 | 1 | US-08-459- Sequence 4, Applicatio  | 7.41e-01 |
| 38 | 66 | 29.3 | 336 | 1 | US-07-667- Sequence 8, Applicatio  | 7.41e-01 |
| 39 | 66 | 29.3 | 509 | 4 | PCI-US95-0 Sequence 8, Applicatio  | 7.41e-01 |
| 40 | 66 | 29.3 | 509 | 2 | US-08-559- Sequence 8, Applicatio  | 7.41e-01 |
| 41 | 66 | 29.3 | 509 | 2 | US-08-749- Sequence 4, Applicatio  | 7.41e-01 |
| 42 | 66 | 29.3 | 615 | 2 | US-08-484- Sequence 38, Applicatio | 7.41e-01 |
| 43 | 65 | 28.9 | 709 | 1 | US-07-814- Sequence 7, Applicatio  | 8.91e-01 |
| 44 | 65 | 28.9 | 709 | 4 | PCI-US92-1 Sequence 7, Applicatio  | 8.91e-01 |
| 45 | 65 | 28.9 | 709 | 1 | US-08-328- Sequence 2, Applicatio  | 8.91e-01 |

A \*GNMENTS

RESULT 1  
ID US-08-452-052-2 STANDARD PRI: 1018 AA.  
XX  
AC XXXXX  
XX

Sequence 2, Application US/08452052

Sequence 2, Application US/08452052

Patent No. 5766922

GENERAL INFORMATION:

APPLICANT: Peles, Elior

TITLE OF INVENTION: FUNCTIONAL LIGANDS FOR THE AXONAL CELL

TITLE OF INVENTION: RECOGNITION MOLECULE CONTACTIN

NUMBER OF SEQUENCES: 2

CORRESPONDENCE ADDRESS:

ADDRESSEE: Peles & Ephraim

STREET: 1155 Avenue of the Americas

CITY: New York

STATE: New York

COUNTRY: U.S.A.

ZIP: 10036-2711

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC Compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0. Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/452,052

FILING DATE: 26-MAY-1995

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Coruzzi, Laura A.

REGISTRATION NUMBER: 30,742

REFERENCE/DOCKET NUMBER: 7683-075

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 790-9090

TELEFAX: (212) 869-8864/9741

TELEX: 66141 PENNIE

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 1018 amino acids

TYPE: amino acid

STRANDEDNESS:

TOPOLOGY: unknown

MOLECULE TYPE: protein

SQ SEQUENCE 101P AA: 113102 MW: 5593224 CN:  
 Query Match 35.1% Score 79: DB 1: Length 1018:  
 Best Local Similarity 34.5% Pred. No. 6,12e-01:  
 Matches 10: Conservative 11: Mismatches 6: Indels 2: Gaps 2:  
 Db 164 FPCSLSYKLNKNEPFIIMKRRFVSCIN 192  
 QY 2 FSKNLNKL N-MPLYAGNKKRFKRVSV 28  
 RESULT 2  
 ID US-08-118-270 76 STANDARD: PRT: 298 AA:  
 XX XXXXXX  
 DE Sequence 76, Application US/08118270  
 CC Sequence 76, Application US/08118270  
 CC Patent No. 550844  
 CC GENERAL INFORMATION:  
 CC APPLICANT: Murphy, Randall B.  
 CC APPLICANT: Schuster, David I.  
 CC TITLE OF INVENTION: POLYPEPTIDES OF G-COUPLED PROTEIN  
 CC TITLE OF INVENTION: RECEPTORS, AND COMPOSITIONS AND METHODS THEREOF  
 CC NUMBER OF SEQUENCES: 348  
 CC CORRESPONDENCE ADDRESS:  
 CC ADDRESSEE: BROWDY AND NEIMARK  
 CC STREET: 419 Seventh Street, N.W., Suite 300  
 CC CITY: Washington  
 CC STATE: D.C.  
 CC COUNTRY: USA  
 CC ZIP: 20004  
 CC COMPUTER READABLE FORM:  
 CC MEDIUM TYPE: Floppy disk  
 CC COMPUTER: IBM PC compatible  
 CC OPERATING SYSTEM: PC-DOS/MS-DOS  
 CC SOFTWARE: Patent In Release #1.0, Version #1.25  
 CC CURRENT APPLICATION DATA:  
 CC APPLICATION NUMBER: US/08/118 270  
 CC FILING DATE: 09-SEP-1993  
 CC PRIOR APPLICATION DATA:  
 CC APPLICATION NUMBER: US 07/943,246  
 CC FILING DATE: 10-SEP-1992  
 CC ATTORNEY/AGENT INFORMATION:  
 CC NAME: Townsend, Kevin G.  
 CC REGISTRATION NUMBER: 34,033  
 CC REFERENCE/DOCKET NUMBER: MURPHY-2A  
 CC TELECOMMUNICATION INFORMATION:  
 CC TELEPHONE: 202-628-5197  
 CC TELEFAX: 202-737-3528  
 CC TELEX: 248633  
 CC INFORMATION FOR SEQ ID NO: 76:  
 CC SEQUENCE CHARACTERISTICS:  
 CC LENGTH: 298 amino acids  
 CC TYPE: amino acid  
 CC STRANDEDNESS: single  
 CC TOPOLOGY: linear  
 CC MOLECULE TYPE: peptide  
 CC SEQUENCE 298 AA: 33548 MW: 467538 CN:  
 Query Match 32.4% Score 73: DB 1: Length 298:  
 Best Local Similarity 32.3% Pred. No. 1.97e-01:  
 Matches 10: Conservative 10: Mismatches 10: Indels 1: Gaps 1:  
 Db 50 LFSKAVIALNMGTF-LGSFPDYVRRVSRIV 79  
 QY 1 LFSKNLNKL N-MPLYAGNKKRFKRVSNVI 31  
 RESULT 4  
 ID US-08-537-002A-3 STANDARD: PRT: 963 AA:  
 XX XXXXXX  
 DE Sequence 3, Application US/08537002A  
 CC Sequence 3, Application US/08537002A  
 CC Patent No. 5773282  
 CC GENERAL INFORMATION:  
 CC APPLICANT: TSUSAKI, Keiji  
 CC APPLICANT: KUBOTA, Michio  
 CC APPLICANT: SUGIMOTO, Toshiyuki

ID PCT-US93-08528-76 STANDARD: PRT: 298 AA:  
 XX XXXXXX  
 DE Sequence 76, Application PCT/US93/08528  
 CC Sequence 76, Application PCT/US93/08528  
 CC GENERAL INFORMATION:  
 CC APPLICANT: New York University  
 CC TITLE OF INVENTION: POLYPEPTIDES OF G-COUPLED PROTEIN  
 CC TITLE OF INVENTION: RECEPTORS, AND COMPOSITIONS AND METHODS THEREOF  
 CC NUMBER OF SEQUENCES: 348  
 CC CORRESPONDENCE ADDRESS:  
 CC ADDRESSEE: BROWDY AND NEIMARK  
 CC STREET: 419 Seventh Street, N.W., Suite 300  
 CC CITY: Washington  
 CC STATE: D.C.  
 CC COUNTRY: USA  
 CC ZIP: 20004  
 CC COMPUTER READABLE FORM:  
 CC MEDIUM TYPE: Floppy disk  
 CC COMPUTER: IBM PC compatible  
 CC OPERATING SYSTEM: PC-DOS/MS-DOS  
 CC SOFTWARE: Patent In Release #1.0, Version #1.25  
 CC CURRENT APPLICATION DATA:  
 CC APPLICATION NUMBER: PCT/US93/08528  
 CC FILING DATE: 09-SEP-1993  
 CC PRIOR APPLICATION DATA:  
 CC APPLICATION NUMBER: US 07/943,246  
 CC FILING DATE: 10-SEP-1992  
 CC ATTORNEY/AGENT INFORMATION:  
 CC NAME: Townsend, Kevin G.  
 CC REGISTRATION NUMBER: 34,033  
 CC REFERENCE/DOCKET NUMBER: MURPHY-2 PCT  
 CC TELECOMMUNICATION INFORMATION:  
 CC TELEPHONE: 202-628-5197  
 CC TELEFAX: 202-737-3528  
 CC TELEX: 248633  
 CC INFORMATION FOR SEQ ID NO: 76:  
 CC SEQUENCE CHARACTERISTICS:  
 CC LENGTH: 298 amino acids  
 CC TYPE: amino acid  
 CC STRANDEDNESS: single  
 CC TOPOLOGY: linear  
 CC MOLECULE TYPE: peptide  
 CC SEQUENCE 298 AA: 33548 MW: 467538 CN:  
 Query Match 32.4% Score 73: DB 4: Length 298:  
 Best Local Similarity 32.3% Pred. No. 1.97e-01:  
 Matches 10: Conservative 10: Mismatches 10: Indels 1: Gaps 1:  
 Db 50 LFSKAVIALNMGTF-LGSFPDYVRRVSRIV 79  
 QY 1 LFSKNLNKL N-MPLYAGNKKRFKRVSNVI 31  
 RESULT 4  
 ID US-08-537-002A-3 STANDARD: PRT: 963 AA:  
 XX XXXXXX  
 DE Sequence 3, Application US/08537002A  
 CC Sequence 3, Application US/08537002A  
 CC Patent No. 5773282  
 CC GENERAL INFORMATION:  
 CC APPLICANT: TSUSAKI, Keiji  
 CC APPLICANT: KUBOTA, Michio  
 CC APPLICANT: SUGIMOTO, Toshiyuki

```

ZIP: 94111
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/58/484,101B
FILING DATE: 07-JUN-1995
CLASSIFICATION: 80C
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/
FILING DATE: 01-JUL-1994
CLASSIFICATION: 80C
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/086,555
FILING DATE: 01-JUL-1993
CLASSIFICATION: 80C
ATTORNEY/AGENT INFORMATION:
NAME: Irecartin, Richard F.
REGISTRATION NUMBER: 31,801
REFERENCE/DOCKET NUMBER: A-575,5-2/RFT
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3349
INFORMATION FOR SEQ ID NO: 48:
SEQUENCE CHARACTERISTICS:
LENGTH: 613 amino acids
TYPE: amio acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE 613 AA: 68332 MW: 1963596 CN:

```

RESULT 6  
ID US-08-484-01B-46 STANDARD PPT: 613 AA  
XX AC XXXXX

AL  
XX  
DT  
XX  
XXXXXX

Sequence 46, Application US/08484101B  
Patent No. 5824868  
GENERAL INFORMATION:  
APPLICANT: California Institute of Technology  
TITLE OF INVENTION: PLANTS HAVING MODIFIED RESPONSE TO  
TITLE OF INVENTION: ETHYLENE  
NUMBER OF SEQUENCES: 50  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Richard F. Treccani  
STREET: 3400 Embarcadero Center, Suite 3400  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/484,101B  
FILING DATE: 07-JUN-1995

CC CLASSIFICATION: 800  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: PCT/US94/  
CC FILING DATE: 01-JUL-1994  
CC CLASSIFICATION: 800  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 08/086,555  
CC FILING DATE: 01-JUL-1993  
CC CLASSIFICATION: 800  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: FERGATIS, Richard F.  
CC REGISTRATION NUMBER: 31,801  
CC REFERENCE/DOCKET NUMBER: A-575,5-2/5FT  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (415) 781-1989  
CC TELEFAX: (415) 398-3249  
CC INFORMATION FOR SEQ ID NO: 46:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 513 amino acids  
CC TYPE: amino acid  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: protein  
CC SEQUENCE: 513 AA: 68312 MW: 1963596 CN:

Query Match: 31.1% Score 70: DB 2: Length 613:  
Best Local Similarity 26.7% Pred. No. 3,49e-01:  
Matches 8: Conservative 11: Mismatches 11: Indels 0: Gaps 0:

DB 439 LSKNLKLNKLPYLTQEARLKKL 468  
QY 2 FSKNLKLNKLPYLTQEARLKKL 3:

RESULT 7  
ID US-08-749-902-6 STANDARD: PRT: 525 AA  
XX  
AC xxxxxx

Sequence 7, Application US/08749902

Sequence 7, Application US/08749902  
Patent No. 5385635  
GENERAL INFORMATION:  
APPLICANT: Bandman, 1a  
APPLICANT: Gold, Sui K.  
APPLICANT: Hillman, Jennifer L.  
TITLE OF INVENTION: NOVEL HUMAN SERINE/THREONINE  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: INCYTE PHARMACEUTICALS, INC.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: US  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/749,902  
FILING DATE: Filed Herewith  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0150 US

TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 525 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: GenBank  
CL NE: 189508  
SQ SEQUENCE 525 AA: 59139 MW: 1455454 CN:

Query Match: 30.7% Score 69: LB 2: Length 525:  
Best Local Similarity 34.6% Pred. No. 4,22e-01:  
Matches 9: Conservative 10: Mismatches 7: Indels 0: Gaps 0:

DB 302 DKILKCKLNKLPYLTQEARLKKL 527  
QY 2 FSKNLKLNKLPYLTQEARLKKL 27

RESULT 8  
ID US-08-749-902-6 STANDARD: PRT: 525 AA  
XX  
AC xxxxxx

Sequence 8, Application US/08749902

Sequence 8, Application US/08749902  
Patent No. 5385635  
GENERAL INFORMATION:  
APPLICANT: Bandman, 1a  
APPLICANT: Gold, Sui K.  
APPLICANT: Hillman, Jennifer L.  
TITLE OF INVENTION: NOVEL HUMAN SERINE/THREONINE  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: INCYTE PHARMACEUTICALS, INC.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: US  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/749,902  
FILING DATE: Filed Herewith  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0150 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 525 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:













\*\*\*\*\*  
[REDACTED]  
\*\*\*\*\*  
(TM)  
\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit,  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Mon Jun 19 16:22:15 2000; Maspar time 15.47 Seconds  
201.982 Million cell updates/sec  
Tabular output not generated.

Title: >US-09-142-524A-4  
Description: (1-31) from US09142524A.pep  
Perfect Score: 225  
Sequence: 1 IFSKNLNKLNMPPLYAGNRRRIKRVSNVI 31

Scoring table: PAM 150  
Gap 11

Searched: 721208 seqs, 100765575 residues  
Post-Processing: Minimum Match 0%  
Listing first 45 summaries

Database: a-pending  
1: PCT 2 06 3 10 60 4 07 5 03 6 08 1 7 08 2 8 08 3 9 08 4 4  
2: US 08 4 13 08 5 12 08 6 13 07 14 08 8 15 08 9 16 09 0 17 09 1  
3: US 09 19 09 3 20 09 4 21 09 5 22 09 6 23 09 7 24 09 8 25 09 9  
25: NEWC9

Statistics: Mean 25.139; Variance 88.013; scale 0.297

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

| SUMMARIES |     |       |  | Pred. No. |  |
|-----------|-----|-------|--|-----------|--|
| Result    | No. | Score | Query Match Length DB ID Description     |           |  |
| 1         | 225 | 100.0 | 31 17 US-09-142- Sequence 4, Applicatio  | 5.32e-15  |  |
| 2         | 223 | 54.7  | 345 3 US-60-109- Sequence 1, Applicatio  | 1.41e-03  |  |
| 3         | 223 | 54.7  | 367 10 US-08-467- Sequence 95, Applicati | 1.41e-03  |  |
| 4         | 223 | 54.7  | 367 10 US-08-468- Sequence 95, Applicati | 1.41e-03  |  |
| 5         | 123 | 54.7  | 367 7 US-08-326- Sequence 95, Applicati  | 1.41e-03  |  |
| 6         | 123 | 54.7  | 367 8 US-08-350- Sequence 95, Applicati  | 1.41e-03  |  |
| 7         | 123 | 54.7  | 367 10 US-08-467- Sequence 95, Applicati | 1.41e-03  |  |
| 8         | 123 | 54.7  | 367 10 US-08-467- Sequence 95, Applicati | 1.41e-03  |  |
| 9         | 123 | 54.7  | 367 10 US-08-467- Sequence 97, Applicati | 1.41e-03  |  |
| 10        | 123 | 54.7  | 370 10 US-08-467- Sequence 97, Applicati | 1.41e-03  |  |
| 11        | 123 | 54.7  | 370 7 US-08-326- Sequence 97, Applicati  | 1.41e-03  |  |
| 12        | 223 | 54.7  | 370 10 US-08-468- Sequence 97, Applicati | 1.41e-03  |  |
| 13        | 223 | 54.7  | 370 10 US-08-467- Sequence 97, Applicati | 1.41e-03  |  |
| 14        | 123 | 54.7  | 370 8 US-08-350- Sequence 97, Applicati  | 1.41e-03  |  |
| 15        | 118 | 52.4  | 60 7 US-08-326- Sequence 62, Applicati   | 4.74e-03  |  |
| 16        | 118 | 52.4  | 60 10 US-08-468- Sequence 62, Applicati  | 4.74e-03  |  |
| 17        | 118 | 52.4  | 60 10 US-08-467- Sequence 62, Applicati  | 4.74e-03  |  |
| 18        | 118 | 52.4  | 60 8 US-08-350- Sequence 62, Applicati   | 4.74e-03  |  |
| 19        | 118 | 52.4  | 60 10 US-08-467- Sequence 62, Applicati  | 4.74e-03  |  |
| 20        | 118 | 52.4  | 60 4 US-07-938- Sequence 62, Applicati   | 4.74e-03  |  |

|    |     |      |        |                 |                         |          |
|----|-----|------|--------|-----------------|-------------------------|----------|
| 21 | 118 | 52.4 | 60 10  | US-09-142-524-4 | Sequence 62, Applicatio | 4.74e-03 |
| 22 | 118 | 52.4 | 134 17 | US-09-142-524-4 | Sequence 3, Applicatio  | 4.74e-03 |
| 23 | 118 | 52.4 | 353 11 | US-08-526-524-4 | Sequence 15, Applicatio | 4.74e-03 |
| 24 | 118 | 52.4 | 374 4  | US-07-938-524-4 | Sequence 2, Applicatio  | 4.74e-03 |
| 25 | 118 | 52.4 | 374 4  | US-07-730-524-4 | Sequence 2, Applicatio  | 4.74e-03 |
| 26 | 118 | 52.4 | 374 8  | US-08-350-524-4 | Sequence 2, Applicatio  | 4.74e-03 |
| 27 | 118 | 52.4 | 374 10 | US-08-467-524-4 | Sequence 2, Applicatio  | 4.74e-03 |
| 28 | 118 | 52.4 | 374 10 | US-08-467-524-4 | Sequence 2, Applicatio  | 4.74e-03 |
| 29 | 118 | 52.4 | 374 4  | US-07-723-524-4 | Sequence 2, Applicatio  | 4.74e-03 |
| 30 | 118 | 52.4 | 374 4  | US-08-467-524-4 | Sequence 2, Applicatio  | 4.74e-03 |
| 31 | 118 | 52.4 | 374 7  | US-08-226-524-4 | Sequence 2, Applicatio  | 4.74e-03 |
| 32 | 118 | 52.4 | 374 10 | US-08-468-524-4 | Sequence 2, Applicatio  | 4.74e-03 |
| 33 | 113 | 50.2 | 20 7   | US-08-226-524-4 | Sequence 33, Applicati  | 1.58e-02 |
| 34 | 113 | 50.2 | 20 10  | US-08-467-524-4 | Sequence 33, Applicati  | 1.58e-02 |
| 35 | 113 | 50.2 | 20 10  | US-08-467-524-4 | Sequence 33, Applicati  | 1.58e-02 |
| 36 | 113 | 50.2 | 20 8   | US-08-350-524-4 | Sequence 33, Applicati  | 1.58e-02 |
| 37 | 113 | 50.2 | 20 10  | US-08-468-524-4 | Sequence 33, Applicati  | 1.58e-02 |
| 38 | 113 | 50.2 | 20 10  | US-08-467-524-4 | Sequence 33, Applicati  | 1.58e-02 |
| 39 | 87  | 38.7 | 105 17 | US-09-142-524-4 | Sequence 5, Applicatio  | 6.51e-01 |
| 40 | 82  | 36.4 | 31 17  | US-09-142-524-4 | Sequence 5, Applicatio  | 1.94e-01 |
| 41 | 80  | 35.6 | 80 17  | US-09-142-524-4 | Sequence 1, Applicatio  | 3.04e-01 |
| 42 | 80  | 35.6 | 377 3  | US-60-096-524-4 | Sequence 15285, Applic  | 3.04e-01 |
| 43 | 80  | 35.6 | 377 18 | US-09-248-524-4 | Sequence 15285, Applic  | 3.04e-01 |
| 44 | 78  | 34.7 | 211 18 | US-09-270-524-4 | Sequence 42685, Applic  | 4.58e-01 |

ALIGNMENTS

RESULT: 1  
ID US-09-142-524-4 STANDARD: PRI: 31 AA.  
AC xxxxxx  
DI  
DT  
XX  
DE Sequence 4, Application US/09142524  
XX Sequence 4, Application US/09142524  
CC GENERAL INFORMATION:  
CC APPLICANT: Sone, Toshio  
CC APPLICANT: Kume, Akinori  
CC APPLICANT: Kairiki, Kazuo  
CC APPLICANT: Iwama, Akiko  
CC APPLICANT: King, Kohsuke  
CC TITLE OF INVENTION: Rept: based immunotherapeutic Agent For Treating  
CC TITLE OF INVENTION: Allergic Diseases  
CC FILE REFERENCE: Patent No. SPO-103  
CC CURRENT APPLICATION NUMBER: US/09/142-524  
CC CURRENT FILING DATE: 1999-01-04  
CC EARLIER APPLICATION NUMBER: B/90/702  
CC EARLIER FILING DATE: 1996-03-10  
CC EARLIER APPLICATION NUMBER: PCT/JF97/00740  
CC EARLIER FILING DATE: 1997-03-10  
CC NUMBER OF SEQ ID NOS: 5  
CC SOFTWARE: Patentin Ver. 2.0  
CC SEQ ID NO 4  
CC LENGTH: 31  
CC TYPE: PRI  
CC ORGANISM: Unknown  
CC FEATURE:  
CC OTHER INFORMATION: Description of Unknown Organism: peptide  
SQ SEQUENCE 31 AA: 3661 MW: 4960 CN:

Query Match 100.0%; core 225; DB 17; Length 31;  
Best Local Similarity 100.0%; red. No. 5.32e-15;  
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 IFSKNLNKLNMPPLYAGNRRRIKRVSNVI 31  
OY 1 IFSKNLNKLNMPPLYAGNRRRIKRVSNVI 31

RESULT 2  
 ID US-60-109-165-1 STANDARD: PRT: 345 AA.  
 XX  
 AC xxxxxx  
 DT  
 XX  
 XX  
 DE  
 XX  
 XX  
 Sequence 1, Application US/60109165  
 Sequence 1, Application US/60109165  
 GENERAL INFORMATION:  
 APPLICANT: GOLDBERG, RANDALL M.  
 APPLICANT: BROOKS, EDWARD G.  
 APPLICANT: MIDORO-HORICHI, TERUMI  
 APPLICANT: KUROSKY, ALEX  
 TITLE OF INVENTION: PROTEIN IN MOUNTAIN CEDAR POLLEN THAT BELONGS TO A  
 TITLE OF INVENTION: FAMILY OF PLANT PATHOGENESIS RELATED PROTEINS  
 FILE REFERENCE: JTSG-230921  
 CURRENT APPLICATION NUMBER: US/60/109,165  
 CURRENT FILING DATE: 1995-11-20  
 NUMBER OF SEQ ID NOS: 5  
 SOFTWARE: Patent In Ver. 2.0  
 SEQ ID NO: 1  
 LENGTH: 346  
 TYPE: PRT  
 ORGANISM: Juniperus ashei  
 SEQUENCE 345 AA: 37639 MW: 619883 CN:  
 Query Match 54.7% Score 123 DB 3 Length 346  
 Best Local Similarity 75.0% Pred. No. 1.41e-03  
 Matches 15: Conservative 4: Mismatches 1: Indels 0: Gaps 0:  
 DB 71 JFSQNNN:KLMPLYVAGHK 90  
 QY 1 JFSKNN:KLMPLYIAGNK 20  
 RESULT 4  
 ID US-08-467-023-95 STANDARD: PRT: 367 AA.  
 XX  
 AC xxxxxx  
 DT  
 XX  
 XX  
 DE  
 XX  
 XX  
 Sequence 95, Application US/08467023  
 Sequence 95, Application US/08467023  
 GENERAL INFORMATION:  
 APPLICANT: Griffiths, Irwin J.  
 APPLICANT: Pollock, Joshua  
 APPLICANT: Bond, Julian P.  
 APPLICANT: Garman, Richard D.  
 APPLICANT: Kuo, Mei-Chang  
 APPLICANT: Young, Steven H.  
 APPLICANT: Brauer, Andrew  
 APPLICANT: Exley, Mark A.  
 APPLICANT: Powers, Steven P.  
 TITLE OF INVENTION: Allergenic Proteins And Peptides From  
 TITLE OF INVENTION: Japanese Cedar Pollen  
 NUMBER OF SEQUENCES: 261  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: ImmuLogic Pharmaceutical Corporation, Inc.  
 STREET: 610 Lincoln St  
 CITY: Waltham  
 STATE: MA  
 COUNTRY: USA  
 ZIP: 02154  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent In Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/468,940  
 FILING DATE:  
 CLASSIFICATION: 424  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 08/350,225  
 FILING DATE: December 6, 1994  
 SOFTWARE: Patent In Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: 08/226,248

CC APPLICATION NUMBER: US/08/467,023  
 CC FILING DATE: June 6, 1995  
 CC CLASSIFICATION: 424  
 CC PRIOR APPLICATION DATA:  
 CC APPLICATION NUMBER: 08/350,225  
 CC FILING DATE: December 6, 1994  
 CC ATTORNEY/AGENT INFORMATION:  
 CC NAME: Jane E. Romilla  
 CC REGISTRATION NUMBER: 38,872  
 CC REFERENCE/DOCKET NUMBER: 025.6 US02 (IM-0280142)  
 CC TELECOMMUNICATION INFORMATION:  
 CC TELEPHONE: (617) 227-7400  
 CC TELEFAX: (617) 227-5941  
 CC INFORMATION FOR SEQ ID NO: 95:  
 CC SEQUENCE CHARACTERISTICS:  
 CC LENGTH: 367 amino acids  
 CC TYPE: amino acid  
 CC TOPOLOGY: linear  
 CC MOLECULE TYPE: protein  
 CC SEQUENCE 367 AA: 39846 MW: 95076 CN:  
 Query Match 54.7% Score 123 DB 3 Length 467  
 Best Local Similarity 75.0% Pred. No. 1.41e-03  
 Matches 15: Conservative 4: Mismatches 1: Indels 0: Gaps 0:  
 DB 92 JFSQNNN:KLMPLYVAGHK 100  
 QY 1 JFSKNN:KLMPLYIAGNK 20  
 RESULT 4  
 ID US-08-468-940-95 STANDARD: PRT: 367 AA.  
 XX  
 AC xxxxxx  
 DT  
 XX  
 XX  
 DE  
 XX  
 XX  
 Sequence 95, Application US/08468940  
 Sequence 95, Application US/08468940  
 GENERAL INFORMATION:  
 APPLICANT: Griffiths, Irwin J.  
 APPLICANT: Pollock, Joshua  
 APPLICANT: Bond, Julian P.  
 APPLICANT: Garman, Richard D.  
 APPLICANT: Kuo, Mei-Chang  
 APPLICANT: Young, Steven H.  
 APPLICANT: Brauer, Andrew  
 APPLICANT: Exley, Mark A.  
 APPLICANT: Powers, Steven P.  
 TITLE OF INVENTION: Allergenic Proteins And Peptides From  
 TITLE OF INVENTION: Japanese Cedar Pollen  
 NUMBER OF SEQUENCES: 261  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: ImmuLogic Pharmaceutical Corporation, Inc.  
 STREET: 610 Lincoln St  
 CITY: Waltham  
 STATE: MA  
 COUNTRY: USA  
 ZIP: 02154  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent In Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/468,940  
 FILING DATE:  
 CLASSIFICATION: 424  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 08/350,225  
 FILING DATE: December 6, 1994  
 SOFTWARE: Patent In Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: 08/226,248

CC FILING DATE: April 8, 1994  
CC APPLICATION NUMBER: 07/938,990  
CC FILING DATE: September 1, 1992  
CC APPLICATION NUMBER: PCT/US93/00139  
CC FILING DATE: January 15, 1993  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Darlene A. Vanstone  
CC REGISTRATION NUMBER: 35,729  
CC REFERENCE/DOCKET NUMBER: 025.6 US (IMI-028CP2)  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (617) 466-6000  
CC TELEFAX: (617) 466-6040  
CC INFORMATION FOR SEQ ID NO: 95:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 367 amino acids  
CC TYPE: amino acid  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: protein  
CC SEQUENCE: 367 AA: 39846 MW: 595078 CN:  
52  
Query Match 54.7% Score 123: DB 10: Length 367:  
Best Local Similarity 75.0% Pred. No. 1.41e-03:  
Matches 15: Conservative 4: Mismatches 1: Indels 0: Gaps 0:  
DB 92 IFSONNMIKLMPLYVASHK 111  
QY 1 IFSKNLNKLNMPYIAGNK 20  
RESULT 5  
ID US-06-226-248A-95 STANDARD: PRI: 367 AA.  
XX  
AC xxxxxx  
XX  
D7  
XX  
DE Sequence 95, Application US/08226248A  
XX  
Sequence 95, Application US/08226248A  
CC GENERAL INFORMATION:  
CC APPLICANT: Griffith, Irwin J.  
CC APPLICANT: Pollock, Joanne  
CC APPLICANT: Bond, Julian F.  
CC APPLICANT: Garman, Richard D.  
CC APPLICANT: Kuo, Mei-Chang  
CC APPLICANT: Yeung, Siu-mei H.  
CC APPLICANT: Brauer, Andrew  
CC APPLICANT: Exley, Mark A.  
CC APPLICANT: Powers, Steven P.  
CC TITLE OF INVENTION: Allergenic Proteins And Peptides From  
CC NUMBER OF SEQUENCES: 201  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Immunologic Pharmaceutical Corporation, Inc.  
CC STREET: 610 Lincoln St  
CC CITY: Waltham  
CC STATE: MA  
CC COUNTRY: USA  
CC ZIP: 02154  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: PatentIn Release #1.0, Version #1.25  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/226,248A  
CC FILING DATE: April 8, 1994  
CC CLASSIFICATION: 424  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: 07/938,990  
CC FILING DATE: September 1, 1992  
CC APPLICATION NUMBER: PCT/US93/00139  
CC FILING DATE: January 15, 1993

CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Darlene A. Vanstone  
CC REGISTRATION NUMBER: 35,729  
CC REFERENCE/DOCKET NUMBER: 025.5 US (IMI-028CP)  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (617) 466-6000  
CC TELEFAX: (617) 466-6040  
CC INFORMATION FOR SEQ ID NO: 95:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 367 amino acids  
CC TYPE: amino acid  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: protein  
CC SEQUENCE: 367 AA: 39846 MW: 595078 CN:  
52  
Query Match 54.7% Score 123: DB 7: Length 367:  
Best Local Similarity 75.0% Pred. No. 1.41e-03:  
Matches 15: Conservative 4: Mismatches 1: Indels 0: Gaps 0:  
DB 92 IFSONNMIKLMPLYVASHK 111  
QY 1 IFSKNLNKLNMPYIAGNK 20  
RESULT 5  
ID US-08-350-225-95 STANDARD: PRI: 367 AA.  
XX  
AC xxxxxx  
XX  
D7  
XX  
DE Sequence 95, Application US/08350225  
XX  
Sequence 95, Application US/08350225  
CC GENERAL INFORMATION:  
CC APPLICANT: Griffith, Irwin J.  
CC APPLICANT: Pollock, Joanne  
CC APPLICANT: Bond, Julian F.  
CC APPLICANT: Garman, Richard D.  
CC APPLICANT: Kuo, Mei-Chang  
CC APPLICANT: Yeung, Siu-mei H.  
CC APPLICANT: Brauer, Andrew  
CC APPLICANT: Exley, Mark A.  
CC APPLICANT: Powers, Steven P.  
CC TITLE OF INVENTION: Allergenic Proteins And Peptides From  
CC NUMBER OF SEQUENCES: 261  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Immunologic Pharmaceutical Corporation, Inc.  
CC STREET: 610 Lincoln St  
CC CITY: Waltham  
CC STATE: MA  
CC COUNTRY: USA  
CC ZIP: 02154  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: PatentIn Release #1.0, Version #1.25  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: U/08/350,225  
CC FILING DATE: December 6, 1994  
CC CLASSIFICATION: 424  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: 08/226,248  
CC FILING DATE: April 8, 1994  
CC APPLICATION NUMBER: 07/938,990  
CC FILING DATE: September 1, 1992  
CC APPLICATION NUMBER: PCT/US93/00139  
CC FILING DATE: January 15, 1993  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Darlene A. Vanstone  
CC REGISTRATION NUMBER: 35,729

CC REFERENCE/DOCKET NUMBER: 025.6 US (IMI-028CP2)  
 CC TELECOMMUNICATION INFORMATION:  
 CC TELEPHONE: (617) 465-6000  
 CC TELEFAX: (617) 466-6040  
 CC INFORMATION FOR SEQ ID NO: 95:  
 CC SEQUENCE CHARACTERISTICS:  
 CC LENGTH: 367 amino acids  
 CC TYPE: amino acid  
 CC TOPOLOGY: linear  
 CC MOLECULE TYPE: protein  
 CC SEQUENCE 367 AA: 39846 MW: 695078 CN:

Query Match 54.7% Score 123; DB 8; Length 367;  
 Best Local Similarity 75.0%; Pred. No. 1.41e-03;  
 Matches 15; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

DB 92 IFSONNLIKLMPLYVAGHK 1-1  
 QY 1 IFSKNLNIKLMPLYAGNK 20

RESULT 7  
 ID US-08-467-006-95 STANDARD: PRT: 367 AA:  
 XX  
 AC xxxxxx

Sequence 95, Application US/08467006  
 GENERAL INFORMATION:  
 APPLICANT: Griffith, Irwin J.  
 APPLICANT: Pollock, Joanne;  
 APPLICANT: Bond, Julian F.  
 APPLICANT: Garman, Richard D.  
 APPLICANT: Kuo, Mel-Chang;  
 APPLICANT: Yeung, Siu-mei H.;  
 APPLICANT: Brauer, Andrew;  
 APPLICANT: Exley, Mark A.;  
 APPLICANT: Powers, Steven P.  
 TITLE OF INVENTION: Allergic Proteins And Filides From  
 TITLE OF INVENTION: Japanese Cedar Pollen  
 NUMBER OF SEQUENCES: 261  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Immunologic Pharmaceutical Corporation, Inc.  
 STREET: 610 Lincoln St  
 CITY: Waltham  
 STATE: MA  
 COUNTRY: USA  
 ZIP: 02154

COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent In Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/467.006  
 FILING DATE: June 6, 1995  
 CLASSIFICATION: 424  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 08/350,225  
 FILING DATE: December 6, 1994  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Jane E. Remillard  
 REGISTRATION NUMBER: 38,872  
 REFERENCE/DOCKET NUMBER: 025.6 USD6 (IMI-028CPD6);  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (617) 227-7400  
 TELEFAX: (617) 227-5941  
 INFORMATION FOR SEQ ID NO: 95:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 367 amino acids

CC TYPE: amino acid  
 CC TOPOLOGY: linear  
 CC MOLECULE TYPE: protein  
 CC SEQUENCE 367 AA: 39846 MW: 695078 CN:  
 Query Match 54.7% Score 123; DB 10; Length 367;  
 Best Local Similarity 75.0%; Pred. No. 1.41e-03;  
 Matches 15; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

DB 92 IFSONNLIKLMPLYVAGHK 1-1  
 QY 1 IFSKNLNIKLMPLYAGNK 20

RESULT 8  
 ID US-08-467-697-95 STANDARD: PRT: 367 AA:  
 XX  
 AC xxxxxx

Sequence 95, Application US/08467697  
 GENERAL INFORMATION:  
 APPLICANT: Griffith, Irwin J.  
 APPLICANT: Pollock, Joanne;  
 APPLICANT: Bond, Julian F.  
 APPLICANT: Garman, Richard D.  
 APPLICANT: Kuo, Mel-Chang;  
 APPLICANT: Yeung, Siu-mei H.;  
 APPLICANT: Brauer, Andrew;  
 APPLICANT: Exley, Mark A.;  
 APPLICANT: Powers, Steven P.  
 TITLE OF INVENTION: Allergic Proteins And Peptides From  
 TITLE OF INVENTION: Japanese Cedar Pollen  
 NUMBER OF SEQUENCES: 261  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Immunologic Pharmaceutical Corporation, Inc.  
 STREET: 610 Lincoln St  
 CITY: Waltham  
 STATE: MA  
 COUNTRY: USA  
 ZIP: 02154

COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent In Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/467.697  
 FILING DATE: June 6, 1995  
 CLASSIFICATION: 424  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 08/350,225  
 FILING DATE: December 6, 1994  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Jane E. Remillard  
 REGISTRATION NUMBER: 38,872  
 REFERENCE/DOCKET NUMBER: 025.6 USD4 (IMI-028CPD4)  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (617) 227-7400  
 TELEFAX: (617) 227-5941  
 INFORMATION FOR SEQ ID NO: 95:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 367 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 CC SEQUENCE 367 AA: 39846 MW: 695078 CN:

Query Match 54.7% Score 123; DB 10; Length 367;  
 Best Local Similarity 75.0%; Pred. No. 1.41e-03;

```

RESULT 10
ID US-08-467-697-97 STANDARD: PRT: 370 AA.
XX
AC xxxxxx
XX
DT
XX
DE
XX
XX
Sequence 97, Application US/08467697
Sequence 97, Application US/08467697
GENERAL INFORMATION:
CC APPLICANT: Griffith, Irwin J.;
CC APPLICANT: Pollock, Anne;
CC APPLICANT: Bond, Julian F.;
CC APPLICANT: Garman, Richard D;
CC APPLICANT: Kuo, Mei-Chang;
CC APPLICANT: Yeung, Siu-mei H.;
CC APPLICANT: Brauer, Andrew;
CC APPLICANT: Exley, Mark A.;
CC APPLICANT: Powers, Steven P.
CC TITLE OF INVENTION: Allergenic Proteins And Peptides From
CC TITLE OF INVENTION: Japanese Cedar Pollen
CC NUMBER OF SEQUENCES: 261
CC CORRESPONDENCE ADDRESS:
CC ADDRESS: Immunologic Pharmaceutical Corporation, Inc.
CC STREET: 610 Lincoln St
CC CITY: Waltham
CC STATE: MA
CC COUNTRY: USA
CC ZIP: 02154
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/467-697
CC FILING DATE: June 6, 1995
CC CLASSIFICATION: 424
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/350,225
CC FILING DATE: December 6, 1994
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Jane E. Remillard
CC REGISTRATION NUMBER: 28,872
CC REFERENCE/DOCKET NUMBER: 025,6 USD4 (IM1-028CPD4)
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (617) 227-7400
CC TELEFAX: (617) 227-5941
CC INFORMATION FOR SEQ ID NO: 97:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 370 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC SEQUENCE 370 AA: 4019; MW: 725506 CN;
SQ
Query Match 54.78; Score 123; DB 10; Length 370;
Best Local Similarity 75.08; Pred.No.1.41e-03;
Matches 15; Conservative 4; Mismatches 1; Indels 0; Gaps 0;
DB 92 IFSONNNIKLKMPLYVAGHK 111
QY 1 IFSKNLNKLNMPLYIAGNK 20
RESULT 11
ID US-08-226-248A-97 STANDARD: PRT: 370 AA.
XX
AC xxxxxx
XX
DT
XX
XX

```

```

DE Sequence 97, Application US/08226248A
XX Sequence 97, Application US/08226248A
CC GENERAL INFORMATION:
CC APPLICANT: Griffith, Irwin J.:
CC APPLICANT: Pollock, Joanne:
CC APPLICANT: Bond, Julian F.:
CC APPLICANT: Garman, Richard D:
CC APPLICANT: Kuo, Mei-Chang:
CC APPLICANT: Yeung, Siu-mei H.:
CC APPLICANT: Brauer, Andrew:
CC APPLICANT: Exley, Mark A.:
CC APPLICANT: Powers, Steven P.:
CC TITLE OF INVENTION: Allergic Proteins And Peptides From
CC TITLE OF INVENTION: Japanese Cedar Pollen
CC NUMBER OF SEQUENCES: 201
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Immunologic Pharmaceutical Corporation, Inc.
CC STREET: 610 Lincoln St
CC CITY: Waltham
CC STATE: MA
CC COUNTRY: USA
CC ZIP: 02154
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patent'n Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/226.248A
CC FILING DATE: April 8, 1994
CC CLASSIFICATION: 424
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/350.225
CC FILING DATE: December 6, 1994
CC APPLICATION NUMBER: 08/226.248
CC FILING DATE: April 8, 1994
CC APPLICATION NUMBER: 07/938.990
CC FILING DATE: September 1, 1992
CC APPLICATION NUMBER: PC/JUS93/00139
CC FILING DATE: January 15, 1993
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Darlene A. Vanstone
CC REGISTRATION NUMBER: 35,729
CC REFERENCE/DOCKET NUMBER: 025.5 US (IM1-028CP)
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (617) 466-6000
CC TELEFAX: (617) 466-6040
CC INFORMATION FOR SEQ ID NO: 97:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 370 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC SEQUENCE 370 AA: 40191 MW: 725506 CN:

Query Match 54.78; Score 123; DB 7; Length 370;
Best Local Similarity 75.08; Pred. No. 1,41e-03;
Matches 15; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

D5 92 IFSQNNIKLKMPLYVAGHK 111
QY 1 IFSKNLNKLNMPLYAGNK 20

RESULT 12
ID US-08-468-940-97 STANDARD: PRT: 370 AA.
XX xxxxxx
XX
XX
XX
XX Sequence 97, Application US/08468940
XX Sequence 97, Application US/08468940
CC GENERAL INFORMATION:
CC APPLICANT: Griffith, Irwin J.:
CC APPLICANT: Pollock, Joanne:
CC APPLICANT: Bond, Julian F.:
CC APPLICANT: Garman, Richard D:
CC APPLICANT: Kuo, Mei-Chang:
CC APPLICANT: Yeung, Siu-mei H.:
CC APPLICANT: Brauer, Andrew:
CC APPLICANT: Exley, Mark A.:
CC APPLICANT: Powers, Steven P.:
CC TITLE OF INVENTION: Allergic Proteins And Peptides From
CC TITLE OF INVENTION: Japanese Cedar Pollen
CC NUMBER OF SEQUENCES: 201
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Immunologic Pharmaceutical Corporation, Inc.
CC STREET: 610 Lincoln St
CC CITY: Waltham
CC STATE: MA
CC COUNTRY: USA
CC ZIP: 02154
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patent'n Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/226.248A
CC FILING DATE: April 8, 1994
CC CLASSIFICATION: 424
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 07/938.990
CC FILING DATE: September 1, 1992
CC APPLICATION NUMBER: PC/JUS93/00139
CC FILING DATE: January 15, 1993
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Darlene A. Vanstone
CC REGISTRATION NUMBER: 35,729
CC REFERENCE/DOCKET NUMBER: 025.5 US (IM1-028CP)
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (617) 466-6000
CC TELEFAX: (617) 466-6040
CC INFORMATION FOR SEQ ID NO: 97:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 370 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC SEQUENCE 370 AA: 40191 MW: 725506 CN:

Query Match 54.78; Score 123; DB 7; Length 370;
Best Local Similarity 75.08; Pred. No. 1,41e-03;
Matches 15; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

D5 92 IFSQNNIKLKMPLYVAGHK 111
QY 1 IFSKNLNKLNMPLYAGNK 20

RESULT 13
ID US-08-467-006-97 STANDARD: PRT: 370 AA.
XX xxxxxx
XX
XX
XX
XX Sequence 97, Application US/08457006
XX Sequence 97, Application US/08457006
CC GENERAL INFORMATION:
CC APPLICANT: Griffith, Irwin J.:
CC APPLICANT: Pollock, Joanne:

```



CC APPLICANT: Bond, Julian F.;  
CC APPLICANT: Garman, Richard D;  
CC APPLICANT: Kuo, Mei-Chang;  
CC APPLICANT: Yeung, Siu-mei H.;  
CC APPLICANT: Brauer, Andrew;  
CC APPLICANT: Exley, Mark A.;  
CC APPLICANT: Powers, Steven P.  
CC TITLE OF INVENTION: Allergenic Proteins And Peptides From  
CC TITLE OF INVENTION: Japanese Cedar Pollen  
CC NUMBER OF SEQUENCES: 261  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: ImmuLogic Pharmaceutical Corporation, Inc.  
CC STREET: 610 Lincoln St  
CC CITY: Waltham  
CC STATE: MA  
CC COUNTRY: USA  
CC ZIP: 02154  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: PatentIn Release #1.0, Version #1.25  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/467,006  
CC FILING DATE: June 6, 1995  
CC CLASSIFICATION: 424  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: 08/350,225  
CC FILING DATE: December 6, 1994  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Case E. Remillard  
CC REGISTRATION NUMBER: 38,872  
CC REFERENCE/DOCKET NUMBER: 025.6 USD6 (IMI-028CPD6)  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (617) 227-7400  
CC TELEFAX: (617) 227-5941  
CC INFORMATION FOR SEQ ID NO: 97:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 370 amino acids  
CC TYPE: amino acid  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: protein  
CC SEQUENCE 370 AA: 40191 MW: 725506 CN:  
Query Match 54.7% Score 123 DB 10: Length 370:  
Best Local Similarity 75.0% Pred. No. 1,41e-03:  
Matches 15: Conservative 4: Mismatches 11: Indels 0: Gaps 0:  
DE 92 IFSONNNIKIKYPLVVAQHK 11  
CY 1 IFSKNNIKLNKPLYIAQNK 20  
RESULT 14  
DE US-08-350-225-97 STANDARD: PRT: 370 AA.  
AC XXXXXX  
XX  
XX  
XX  
XX  
DE Sequence 97, Application US/08350225  
CC Sequence 97, Application US/08350225  
CC GENERAL INFORMATION:  
CC APPLICANT: Griffith, Irwin J.;  
CC APPLICANT: Pollock, Joanne;  
CC APPLICANT: Bond, Julian F.;  
CC APPLICANT: Garman, Richard D;  
CC APPLICANT: Kuo, Mei-Chang;  
CC APPLICANT: Yeung, Siu-mei H.;  
CC APPLICANT: Brauer, Andrew;  
CC APPLICANT: Exley, Mark A.;  
CC APPLICANT: Powers, Steven P.

CC TITLE OF INVENTION: Allergenic Proteins And Peptides From  
CC TITLE OF INVENTION: Japanese Cedar Pollen  
CC NUMBER OF SEQUENCES: 26;  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: ImmuLogic Pharmaceutical Corporation, Inc.  
CC STREET: 610 Lincoln S  
CC CITY: Waltham  
CC STATE: MA  
CC COUNTRY: USA  
CC ZIP: 02154  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: PatentIn Release #1.0, Version #1.25  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/350,225  
CC FILING DATE: December 6, 1994  
CC CLASSIFICATION: 424  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: 07/226,248  
CC FILING DATE: April 6, 1994  
CC APPLICATION NUMBER: 07/938,990  
CC FILING DATE: September 1, 1992  
CC APPLICATION NUMBER: PCT/US93/00139  
CC FILING DATE: January 1, 1993  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Darlene A. Vanstine  
CC REGISTRATION NUMBER: 3,729  
CC REFERENCE/DOCKET NUMBER: 025.6 US (IMI-028CP2)  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (617) 466-6300  
CC TELEFAX: (617) 466-6300  
CC INFORMATION FOR SEQ ID NO: 97:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 370 amino acids  
CC TYPE: amino acid  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: protein  
CC SEQUENCE 370 AA: 40191 MW: 725506 CN:  
Query Match 54.7% Score 123 DB 8: Length 370:  
Best Local Similarity 75.0% Pred. No. 1,41e-03:  
Matches 15: Conservative 4: Mismatches 11: Indels 0: Gaps 0:  
DE 92 IFSONNNIKIKYPLVVAQHK 11  
CY 1 IFSKNNIKLNKPLYIAQNK 20  
RESULT 15  
DE US-08-226-248A-62 STANDARD: PRT: 60 AA.  
AC XXXXXX  
XX  
XX  
XX  
XX  
DE Sequence 62, Application US/08226248A  
CC Sequence 62, Application US/08226248A  
CC GENERAL INFORMATION:  
CC APPLICANT: Griffith, Irwin J.;  
CC APPLICANT: Pollock, Joanne;  
CC APPLICANT: Bond, Julian F.;  
CC APPLICANT: Garman, Richard D;  
CC APPLICANT: Kuo, Mei-Chang;  
CC APPLICANT: Yeung, Siu-mei H.;  
CC APPLICANT: Brauer, Andrew;  
CC APPLICANT: Exley, Mark A.;  
CC APPLICANT: Powers, Steven P.  
CC TITLE OF INVENTION: Allergenic Proteins And Peptides From  
CC TITLE OF INVENTION: Japanese Cedar Pollen  
CC NUMBER OF SEQUENCES: 201

```

CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Immunologic Pharmaceutical Corporation, Inc.
CC STREET: 610 Lincoln St
CC CITY: Waltham
CC STATE: MA
CC COUNTRY: USA
CC ZIP: 02154
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/225,248A
CC FILING DATE: April 8, 1994
CC CLASSIFICATION: 424
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 07/938,990
CC FILING DATE: September 1, 1992
CC APPLICATION NUMBER: PCT/US93/00139
CC FILING DATE: January 15, 1993
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Darlene A. VanStone
CC REGISTRATION NUMBER: 35,729
CC REFERENCE/DOCKET NUMBER: 925.5 US (MI-028CP)
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (617) 465-6000
CC TELEFAX: (617) 465-6040
CC INFORMATION FOR SEQ ID NO: 62:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 60 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC FRAGMENT TYPE: internal
CC SEQUENCE 60 AA: 6644 MW: 19464 CN:

```

```

Query Match: 52.4% Score 18: 18 7: length 60:
Best Local Similarity 60.0% Pred. No. 4.7 e-03:
Matches 18: Conservative 4: Mismatch 7: Gaps 1:

```

```

DB 11 IFSONMNIKMPMYIAGYKT-FDGRCAQV 39
   11111111111111111111111111111111
QY 1 FSKNLNLIKMPLYIAGNKAER-KRVSNV 30

```

```

Search completed: Mon Jun 19 10:22:45 2000
JCL time : 30 secs

```

\*\*\*\*\*  
[W][E][B][S][I][T][E]  
[I][N][F][O]  
[F][O][R]  
[M][O][R][E]  
[I][N][F][O]  
[O][N]  
[T][H][E]  
[I][N][T][E][R]  
[N][E][T]  
[A][N][D]  
[O][T][H][E]  
[R]  
[I][N]  
[F][O]  
[R]  
[M]  
[O]  
R  
(TM)  
\*\*\*\*\*

Release 3.1A John F. Collins, BioComputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd  
MPatch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Mon Jun 19 16:00:33 2000: MasPar time 6.43 Seconds  
tabular output not generated. 227.254 Million cell updates/sec

Title: >US-09-142-524A-4  
Description: (1-31) from US09142524A.pep  
Perfect Score: 225  
Sequence: 1 IFSKNLNKLNPLYIAGNKRRIKRVSNV: 31  
Scoring table: PAM 150  
Gap 11  
Searched: 142580 seqs, 47172406 residues  
Post-processing: Minimum Match 0%  
Listing first 45 summaries  
Database: pir62  
1:pir1 2:pir2 3:pir3 4:pir4  
Statistics: Mean 33.568; Variance 61.790; scale 0.545

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.  
SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description | Pred. No.   |
|------------|-------|-------------|--------|-------|-------------|---|
| 1          | 118   | 52.4        | 374    | 2     | JC2124      | major allergen Cry j 1 precursor (clone pCCi-15) - Japanese cedar |
| 2          | 118   | 52.4        | 374    | 2     | JC2123      | major allergen Cry j 1 precursor (clone pCCi-15) - Japanese cedar |
| 3          | 84    | 37.3        | 741    | 2     | T15992      | ethylene receptor hom   |
| 4          | 83    | 36.9        | 898    | 2     | S2826       | hypothetical protein  |
| 5          | 81    | 36.0        | 177    | 2     | S3852       | ribosomal protein   |
| 6          | 81    | 36.0        | 397    | 2     | S26211      | pectate lyase EC 4.2  |
| 7          | 81    | 36.0        | 460    | 1     | E70132      | hypothetical protein  |
| 8          | 81    | 36.0        | 3140   | 1     | GNVSR4      | genome polyprotein  |
| 9          | 81    | 36.0        | 3141   | 1     | GNVSPD      | genome polyprotein  |
| 10         | 78    | 34.7        | 143    | 2     | S45433      | hypothetical protein  |
| 11         | 77    | 34.2        | 480    | 2     | B70367      | hypothetical protein  |
| 12         | 76    | 33.8        | 532    | 2     | T16752      | hypothetical protein  |
| 13         | 76    | 33.8        | 569    | 2     | E64427      | hypothetical protein  |
| 14         | 76    | 33.8        | 3140   | 2     | S47508      | genome polyprotein  |
| 15         | 75    | 33.3        | 273    | 2     | H64420      | hypothetical protein  |
| 16         | 75    | 33.3        | 347    | 2     | S59837      | probable membrane pro   |
| 17         | 74    | 32.9        | 90     | 2     | JQ1098      | auxin-induced protein   |
| 18         | 74    | 32.9        | 92     | 2     | T10942      | auxin-induced protein   |
| 19         | 74    | 32.9        | 189    | 2     | E64371      | hypothetical protein  |
| 20         | 74    | 32.9        | 331    | 2     | S26765      | hypothetical protein  |
| 21         | 74    | 32.9        | 393    | 2     | F75130      | hypothetical protein  |
| 22         | 73    | 32.4        | 343    | 2     | A35639      | G protein-coupled rec   |
| 23         | 73    | 32.4        | 3125   | 1     | GNVSPD      | genome polyprotein  |

|    |    |      |      |   |        |                                     |
|----|----|------|------|---|--------|-------------------------------------|
| 24 | 72 | 32.0 | 113  | 2 | T01657 | hypothetical protein                |
| 25 | 72 | 32.0 | 238  | 2 | S50809 | ribosomal protein L17               |
| 26 | 72 | 32.0 | 309  | 2 | S69201 | invE protein - Shigella dysenteriae |
| 27 | 72 | 32.0 | 309  | 2 | A43859 | IPase - Shigella dysenteriae        |
| 28 | 72 | 32.0 | 309  | 2 | S04379 | virB protein - Shigella dysenteriae |
| 29 | 72 | 32.0 | 327  | 2 | T09027 | probable zinc finger                |
| 30 | 71 | 31.6 | 67   | 2 | S20967 | hypothetical protein                |
| 31 | 71 | 31.6 | 136  | 2 | R5BV51 | ribosomal protein S17               |
| 32 | 71 | 31.6 | 177  | 2 | A75104 | hypothetical protein                |
| 33 | 71 | 31.6 | 255  | 2 | B64175 | repressor modE homolog              |
| 34 | 71 | 31.6 | 845  | 1 | C73079 | endopeptidase CifAip                |
| 35 | 71 | 31.6 | 854  | 1 | C71533 | endopeptidase CifAip                |
| 36 | 71 | 31.6 | 915  | 1 | WMREV3 | ribonucleoside-diphosph             |
| 37 | 70 | 31.1 | 472  | 2 | E70325 | hypothetical protein                |
| 38 | 70 | 31.1 | 613  | 2 | T14432 | probable ethylene rec               |
| 39 | 70 | 31.1 | 613  | 2 | T00758 | ethylene response sen               |
| 40 | 70 | 31.1 | 618  | 2 | G70350 | conserved hypothetical              |
| 41 | 70 | 31.1 | 746  | 2 | S06049 | probable GTP-glycerol               |
| 42 | 70 | 31.1 | 750  | 2 | T04441 | hypothetical protein                |
| 43 | 70 | 31.1 | 1518 | 2 | PQ0221 | polyprotein - plum po               |
| 44 | 70 | 31.1 | 1533 | 2 | F71274 | hypothetical protein                |
| 45 | 67 | 30.7 | 525  | 1 | A41687 | ribosomal protein S6                |

ALGNMENTS

|                      |   |   |                |
|----------------------|---|---|----------------|
| RESULT               | 1 | JC2124  | *type complete |
| ENTRY                |   | major allergen Cry j 1 precursor (clone pCCi-15) - Japanese cedar                             |                |
| TITLE                |   |   |                |
| ORGANISM             |   | *formal_name Cryptomeria japonica *common_name Japanese cedar                                 |                |
| DATE                 |   | 14-Jul-1994 *sequence_revision 14-Jul-1994 *text_change 26-Aug-1999                           |                |
| ACCESSIONS           |   | JC2124  |                |
| REFERENCE            |   | JC2123  |                |
| *authors             |   | Sone, T.; Komiya, N.; Shimizu, K.; Kusakabe, T.; Morikubo, K.; Kino, K.                       |                |
| *journal             |   | Biochem. Biophys. Res. Commun. (1994) 199:619-625   |                |
| *title               |   | Cloning and sequencing of cDNA coding for Cry j 1, a major allergen of Japanese cedar pollen. |                |
| *cross-references    |   | NCBI:94183234   |                |
| *accession           |   | JC2124  |                |
| *molecule_type       |   | mRNA  |                |
| *residues            |   | 1-374   |                |
| *cross-references    |   | GI:226545; NID:433633; PID:3100687; PID:3349414   |                |
| *experimental_source |   | pollen  |                |
| *note                |   | the authors described carbohydrate binding site for residue 279                               |                |

CLASSIFICATION

|                 |  |  |  |
|-----------------|--|--|--|
| NETWORKS        |  | *superfamily family 1474   |  |
| FEATURE         |  | glycoprotein, pollen   |  |
| 1-21            |  |  |  |
| 22-374          |  |  |  |
| 158-191,293,354 |  | *domain signal sequence *status predicted *label NID                         |  |
|                 |  | *product major allergen Cry j 1 (clone pCCi-15) *status predicted *label NID |  |
|                 |  | *binding-site carbohydrate (Asn) (covalent) *status predicted                |  |
| SUMMARY         |  | *length 374 *molecular-weight 46702 *checksum 3692                           |  |

Query Match 52.4%; Score 118; DB 2; Length 374;  
Best Local Similarity 60.0%; Pred. No. 555e-07;  
Matches 18; Conservative 4; Mismatches 7; Indels 1; Gaps 1;

|    |    |                                  |  |
|----|----|----------------------------------|--|
| DB | 92 | IFSNNNIKMKPMYIAGYKT:FDGRGAOV 120 |  |
| QY | 1  | IFSNNNIKLNPLYIAGNKRRIKRVSNV 30   |  |
|    |    | :     :                          |  |

RESULT 2

|          |  |                |
|----------|--|----------------|
| ENTRY    | JC2123   | *type complete |
| TITLE    | major allergen Cry j 1 precursor (clone pCCi-2-2) - Japanese cedar |                |
| ORGANISM | *formal_name Cryptomeria japonica *common_name Japanese cedar      |                |
| DATE     | 14-Jul-1994 *sequence_revision 14-Jul-1994 *text_change            |                |

```

ACCESSIONS      26-Aug-1999
REFERENCE        J02123: PC2065
AUTHORS          Sone, T.; Komiyama, N.; Shimizu, K.; Sakabe, T.; Morikubo,
                  K.; Kinc, K.
JOURNAL          Biochem. Biophys. Res. Commun. (1994) 199:610-625
TITLE            Cloning and sequencing of cDNA coding for Cry j I, a major
                  allergen of Japanese cedar pollen.
CROSS-REFERENCES MIMD:9419334
ACCESSION        J02123
MOLECULE_TYPE    mRNA
RESIDUES         1-374 ##label: SON
CROSS-REFERENCES GB:025544; MID:q49363; PID:d1006086; PID:q493632
EXPERIMENTAL_SOURCE pollen
ACCESSION        PC2065
MOLECULE_TYPE    protein
RESIDUES         22-53;58-81;219-232;236-258;269-307;346-372 ##label: S02
NOTE             The authors described carbohydrate binding site for
                  residue 279
CLASSIFICATION   #superfamily: pectate lyase LA759
KEYWORDS         glycoprotein; pollen
FEATURE          1-21 #domain signal: sequence #status: predicted #label: SIGN
                  22-374 #product: major allergen Cry j 1: (clone PC02-2-2) #status:
                  predicted #label: RAIN
                  158,191,203,354 #binding_site: carbohydrate (Asn) (covalent) #status:
                  predicted
SUMMARY          #length: 374 #molecular_weight: 43645 #checksum: 2920
Query Match     52.4% Score 118; DB 2; Length 374;
Best Local Similarity 60.0%; Pred. NO. 5.55e-07;
Matches 18; Conservative 4; Mismatches 7; Indels 1; Gaps 1;

DB 92 IFSCNNMILKMPYIAGYKT-FDGRGAV 120
QY 1 IFSKNNILKMLPLYIAGNKRRIKRVSNV 30

RESULT 3
ENTRY   #type: complete
TITLE   ethylene receptor homolog - apple tree
ORGANISM #normal_name: Malus domestica #common_name: apple tree
DATE    20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change
        20-Sep-1999
ACCESSIONS  F16992
REFERENCE   Z16395
AUTHORS     Chen, H.H.; Cherng, Y.Y.; Yang, S.F.; Shaw, J.F.
JOURNAL     Plant Physiology (1998) 117:1125
TITLE       Isolation and characterization of a broccoli cDNA (Accession
            No. AF047477) encoding an ERS-type ethylene receptor
            (PGR98-123).
ACCESSION   F16992
STATUS      Preliminary; translated from GB/EMBL/DBJ
MOLECULE_TYPE mRNA
RESIDUES    1-741 ##label: LEE
CROSS-REFERENCES EMBL:AF032448; MID:q341050; PID:q3411051
EXPERIMENTAL_SOURCE cultivar Granny Smith; ripening fruit
GENETICS
NOTE        F16992
SUMMARY     #length 741; #molecular_weight 82967 #checksum 4504
Query Match   37.3% Score 84; DB 2; Length 741;
Best Local Similarity 33.3%; Pred. NO. 1.97e-01;
Matches 10; Conservative 10; Mismatches 10; Indels 0; Gaps 0;

DB 439 LSVTLNIAADLPYIAIDKRLMOTILNV 468
QY 2 IFKNNILKMLPLYIAGNKRRIKRVSNV 31

RESULT 4
ENTRY   #type: complete
TITLE   hypothetical protein YMR066w - yeast (Saccharomyces

```

```

cerevisiae)
ALTERNATE_NAMES hypothetical protein YMR916.05
ORGANISM          #formal_name: Saccharomyces cerevisiae
DATE              19-May-1995 #sequence_revision 01-Sep-1995 #text_change:
                  21-Nov-1997
ACCESSIONS        S52826
REFERENCE          S52814
AUTHORS           Pearson, D.; Bowman, S.
SUBMISSION         Submitted to the EMBL Data Library, April 1995
ACCESSION          S52826
MOLECULE_TYPE      #molecule_type: DNA
RESIDUES           1-898 ##label: PEA
CROSS-REFERENCES  EMBL:Z48952; MID:q763008; PID:q763013; MIPS:YEP_PEA
EXPERIMENTAL_SOURCE strain AB972
GENETICS
FEATURE          138 #map_position: 138
SUMMARY          #length 898 #molecular_weight: 104747 #checksum 7641
Query Match     36.9% Score 83; DB 2; Length 898;
Best Local Similarity 37.0%; Pred. NO. 2.78e-01;
Matches 10; Conservative 11; Mismatches 5; Indels 1; Gaps 1;

DB 785 DLNIDYDLSMYLKGKRCYLROISNT 811
QY 5 NLNKLNMPLYIA-GNKRRIKRVSNV 30

RESULT 5
ENTRY   #type: complete
TITLE   ribosomal protein L5 - Acanthamoeba castellanii (Pilot edition
        (SGC5)
ORGANISM #formal_name: mitochondrion Acanthamoeba castellanii
DATE     15-Jul-1995 #sequence_revision 01-Sep-1995 #text_change:
        26-Feb-1999
ACCESSIONS  S53852
REFERENCE   S53825
AUTHORS     Burger, G.; Piller, L.; Loneragan, K.M.; Gray, M.W.
JOURNAL     J. Mol. Biol. (1995) 245:522-537
TITLE       The mitochondrial DNA of the amoeboid protozoan, Acanthamoeba
        castellanii: complete sequence, gene content and amino
        acid sequence organization.
CROSS-REFERENCES MIMD:95147275
ACCESSION      S53852
STATUS         nucleic acid sequence not shown; translation not shown
MOLECULE_TYPE  DNA
RESIDUES       1-177 ##label: RRP
CROSS-REFERENCES GB:012894; MID:q6624; PID:q6624
EXPERIMENTAL_SOURCE strain Acanthamoeba castellanii
NOTE           the nucleotide sequence was submitted to the EMBL Data
        Library, July 1994
GENETICS
NOTE          mitochondrion
GENETIC_CODE   SGC6
KEYWORDS       mitochondrion
SUMMARY        #length 177; #molecular_weight 21764 #checksum 4772
Query Match     36.0% Score 81; DB 2; Length 177;
Best Local Similarity 50.0%; Pred. NO. 5.49e-01;
Matches 13; Conservative 6; Mismatches 4; Indels 3; Gaps 3;

DB 46 RSLKPKIKLYLWLYIIS-NQKPKRIKV 70
QY 4 KNLN-IKLN-WPLYIAGNKRRIKRV 27

RESULT 6
ENTRY   #type: complete
TITLE   pectate lyase (EC 4.2.2.2) - common tobacco
ORGANISM #formal_name: Nicotiana tabacum #common_name: common tobacco
DATE     12-Feb-1993 #sequence_revision 12-Feb-1993 #text_change:
        29-Sep-1999
ACCESSIONS      S26211; S26212; S21933; S22753; S22754
REFERENCE        S26211; S26212

```

```

#authors Rogers, H.J.; Harvey, A.; Lonsdale, D.M.
#journal Plant Mol. Biol. (1992) 20:493-502
#title Isolation and characterization of a tobacco gene with
#       homology to pectate lyase which is specifically expressed
#       during microsporogenesis.
#cross-references M01D:93043039
#accession S2621
#molecule-type DNA
#residues 1-397 #label ROG
#cross-references EMBL:X57158; NID:g19907; PIDN:CAA7630.1; PID:g19908
#accession S2621
#molecule-type mRNA
#residues 119-255, 'C', 157-188, 190, 'G', 191-199, 'D', 200, 'R', 203-248,
#       'N', 250-381 #label RO2
#cross-references EMBL:X67159
#note Translation of the nucleotide sequence is not complete
#accession S21933
#authors Lonsdale, D.M.
#submission submitted to the EMBL Data Library, July 1991
#accession S21933
#molecule-type DNA
#residues 1-397 #label LON
#cross-references EMBL:X61102; NID:g19961; PIDN:CAA434.4.1; P.D:g19982
GENETICS
#introns 193/1; 293/2
CLASSIFICATION
#superfamily pectate lyase LAT59
#carbon-oxygen lyase
KEYWORDS
#length 397 #molecular-weight 44351 #checksum 1480
Query Match 36.0% Score 81; DB 2; Length 397;
Best Local Similarity 50.0%; Pred. No. 5.49e-01;
Matches 10; Conservative 7; Mismatches 3; Indels 0; Gaps 0;
Db 116 IFGKNMARIK:SRRLIV:SNK 135
QY 1 FSKNLRN:KLNMPYIAGNK 20

RESULT 7
ENTRY E70132 #type complete
#title Hypothetical protein BB0261 - Lyme disease spirochete
#       spirochete
#organism Borrelia burgdorferi #common_name Lyme disease
#date 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change
#accession E70132
#authors Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.;
#       Clayton, R.; Lathigra, R.; White, O.; Ketchum, K.A.;
#       Dodson, R.; Hickey, E.K.; Gwinn, M.; Dougherty, B.; Tomb,
#       J.F.; Fleischmann, R.D.; Richardson, D.; Peterson, J.;
#       Keravane, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.;
#       Vugt, R.V.; Palmer, N.; Adams, M.D.; Gocayne, J.; Weidman,
#       J.; Utterback, T.; Watthey, L.; McDonald, L.; Artach, P.;
#       Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst,
#       K.; Roberts, K.; Hatch, B.; Smith, H.O.; Venter, J.C.
#journal Nature (1997) 350:580-586
#title Genomic sequence of a Lyme disease spirochete, Borrelia
#       burgdorferi.
#cross-references M01D:98065943
#accession E70132
#status preliminary; nucleic acid sequence not shown;
#       translation not shown
#molecule-type DNA
#residues 1-460 #label KLE
#cross-references GB:AE001137; GB:AE000783; NID:g2688160; PID:g2688209;
#       TIGR:BB0261
#experimental_source strain B31
CLASSIFICATION
#superfamily hypothetical protein BA0261; tetratricopeptide
#       repeat homology
FEATURE
#domain tetratricopeptide repeat homology #label TT1\
#domain tetratricopeptide repeat homology #label TT2\

```

```

205-238 #domain tetratricopeptide repeat homology #label TT3\
239-272 #domain tetratricopeptide repeat homology #label TT4\
278-310 #domain tetratricopeptide repeat homology #status
#       atypical #label T15\
311-344 #domain tetratricopeptide repeat homology #label TT5
SUMMARY
#length 460 #molecular-weight 54024 #checksum 9109
Query Match 36.0% Score 81; DB 1; Length 460;
Best Local Similarity 33.3%; Pred. No. 5.49e-01;
Matches 10; Conservative 10; Mismatches 10; Indels 0; Gaps 0;
Db 320 FSKRLNENIALRYLRSSKENEKIANE11 349
QY 2 FSKNLRN:KLNMPYIAGNRRFIKRVSNVI 31

RESULT 8
ENTRY GNV5RA #type complete
#title genome polyprotein - plum pox virus (strain Raskovic)
#contains 29K protein; 34K protein; Coat protein; cylindrical inclusion
#       protein; helper-component protein; nuclear inclusion
#       protein a; nuclear inclusion protein b
#organism host Nicotiana glauca (tobacco)
#note 30-Jun-1993 #sequence_revision 30-Sep-1993 #text_change
#date 23-Jul-1999
#accessions A60009; A42761; B42761; JA0138
#reference A65009
#authors Lain, S.; Riechmann, J.L.; Garcia, J.A.
#journal Virus Res. (1989) 13:157-172
#title The complete nucleotide sequence of plum pox polyvirus RNA.
#cross-references M01D:89370814
#accession A60009
#molecule-type genomic RNA
#residues 1-3140 #label LAT
#reference A42761
#authors Lain, S.; Riechmann, J.L.; Mendez, E.; Garcia, J.A.
#journal Virus Res. (1988) 10:325-342
#title Nucleotide sequence of the 3' terminal region of plum pox
#       polyvirus RNA.
#accession A42761
#molecule-type genomic RN\
#residues 2263-2874; B 2876-3140 #label LAT
#cross-references GB:M21817; NID:g333301; PIDN:AAA85458.1; PID:g333302
#accession B42761
#molecule-type protein.
#residues 2811-2815; 2881-2893 #label LA2
CLASSIFICATION
#superfamily tobacco etch virus genome polyprotein; GEA0/H
#box helicas homology
KEYWORDS
ATP; coat pro. no cylindrical inclusion protein;
genome-linked protein; inclusion protein; nucleus; p-loop;
phosphoprotein; polyprotein
FEATURE
#product 34K protein #status predicted #label PRV\
#product helper component protein #status predicted
#label HCP\
#product 29K protein #status predicted #label PRV\
#product cylindrical inclusion protein #status predicted
#label CIP\
#domain DEAD/H box helicase homology #label DEAD\
#region nucleotide-binding motif A (p-loop)\
#region nucleotide-binding motif B\
#region DEXH motif\
#product nuclear inclusion protein a #status predicted
#label NIA\
#product nuclear inclusion protein b #status predicted
#label NIB\
#product coat protein #status predicted #label CCP\
#binding-site phosphoryl-RNA (Tyr) (covalent) #status
#       predicted
SUMMARY
#length 3140 #molecular-weight 355579 #checksum 7922
Query Match 36.0% Score 81; DB 1; Length 3140;

```

```

Best Local Similarity 45.7%: Pred No. 5,49e-01:
Matches 14: Conservative 8: Mismatches 5: Indels 3: Gaps 3:

Db 2410 IFS-ALNKAAGVAGLY-SGKKRDYFKNVSD 2437
11 1111 1111 1111 1111
QY 1 IFSKNLNKLNN-PLYIAGNKRREIKRVSN 29

RESULT 9
ENTRY genome polyprotein - plum pox virus (strain D)
CONTAINS 29k protein; 34k protein; coat protein; cy:indirect inclusion
protein; helper component protein; nuclear inclusion
protein a; nuclear inclusion protein b
ORGANISM *formal name plum pox virus, PPV
DATE 30-Sep-1991 *sequence_revision 30-Sep-1991 *text_change
05-Jun-1998
ACCESSIONS S06929
REFERENCE S06929
AUTHORS Teycheney, P.V.; Taverl, G.; Deibos, R.; Ravelonandro, M.;
Cunee, J.
JOURNAL Nucleic Acids Res. (1989) 17:10115-10116
TITLE The complete nucleotide sequence of plum pox virus RNA
(strain D)
CROSS-REFERENCES M01D:SC098790
ACCESSION S06929
STATUS translation not shown
MOLECULE_TYPE genomic RNA
RESIDUES 1-341 *label TEY
CROSS-REFERENCES EMBL:X16415
CLASSIFICATION superfamily tobacco etch virus genome polyprotein: DEAD/H
box helicase homology
KEYWORDS Arp; coat protein; cy:indirect inclusion protein;
genome-linked protein; inclusion protein; nucleus: p-loop;
phosphoprotein; polyprotein
FEATURE
1-302
302-915
916-1169
1170-1857
1254-1527
1258-1261
1339-1344
1343-1346
1858-2233
2294-2811
2812-3141
1920
SUMMARY *length 3141 *molecular_weight 35555 *checksum 3143
Query Match 36.0%: Score 81: DB 1: Length 3141:
Best Local Similarity 45.7%: Pred No. 5,49e-01:
Matches 14: Conservative 8: Mismatches 5: Indels 3: Gaps 3:

Db 2411 IFS-ALNKAAGVAGLY-SGKKRDYFKNVSD 2438
11 1111 1111 1111 1111
QY 1 IFSKNLNKLNN-PLYIAGNKRREIKRVSN 29

RESULT 10
ENTRY genome polyprotein - yeast (Saccharomyces
cerevisiae)
CONTAINS hypothetical protein YAL077w - yeast (Saccharomyces
cerevisiae)
ORGANISM *formal name Saccharomyces cerevisiae
DATE 09-Jun-1994 *sequence_revision 09-Sep-1994 *text_change
21-Nov-1997
ACCESSIONS S45433
REFERENCE S45813; S45818; S59230

```

```

REFERENCE S45387
AUTHORS Obermaier, B.; Gassenhuber, J.; Piravandi, E.; Domdey, H.
SUBMISSION Submitted to the EMBL Data Library, May 1994
DESCRIPTION Sequence analysis of a 78.6 kb segment of the left end of
Saccharomyces cerevisiae chromosome III.
ACCESSION S45433
MOLECULE_TYPE DNA
RESIDUES 1-143 *label OBE
CROSS-REFERENCES EMBL:X79489: NID:q45661: PID:q45670w
EXPERIMENTAL_SOURCE strain S288C
REFERENCE S45813
AUTHORS Contreras, R.; Liu, W.; Logghe, M.; Moremans, F.
SUBMISSION Submitted to the Protein Sequence Database, August 1994
ACCESSION S45813
MOLECULE_TYPE DNA
RESIDUES 1-143 *label N
CROSS-REFERENCES EMBL:2355: NID:q59612: PID:q596122: MIPS:YML077w
EXPERIMENTAL_SOURCE strain S8C
REFERENCE S45815
AUTHORS Domdey, H.; Gassenhuber, H.; Obermaier, B.; Piravandi, E.
SUBMISSION Submitted to the Protein Sequence Database, August 1994
ACCESSION S45818
MOLECULE_TYPE DNA
RESIDUES 1-143 *label DOM
CROSS-REFERENCES EMBL:235838: NID:q59612: PID:q596122: MIPS:YML077w
EXPERIMENTAL_SOURCE strain S288C
REFERENCE S59184
AUTHORS Obermaier, B.; Gassenhuber, J.; Piravandi, E.; Domdey, H.
JOURNAL Yeast (1995) 11:1103-1112
TITLE Sequence analysis of a 78.6 kb segment of the left end of
Saccharomyces cerevisiae chromosome III.
CROSS-REFERENCES M01D:96076635
ACCESSION S59230
STATUS nucleic acid sequence not shown: translation not shown
MOLECULE_TYPE DNA
RESIDUES 1-143 *label OMW
CROSS-REFERENCES EMBL:X79489: NID:q45661: PID:q45670c
NOTE the nucleotide sequence was submitted to the EMBL Data
Library, May 1994
GENETICS
MAP_POSITION 21
MAP_LENGTH 143 *molecular_weight 16327 *checksum 2447
SUMMARY
Query Match 34.7%: Score 78: DB 2: Length 143:
Best Local Similarity 33.3%: Pred No. 1,50e+00:
Matches 7: Conservative 1: Mismatches 2: Indels 0: Gaps 0:

Db 25 NLNKN:HLVYITSSREKIVR 4-
11111111111111111111
QY 5 NLNKLNNPPLYAGNKRREIKR 25

RESULT 11
ENTRY hypothetical protein aq765 - Aquifex aeolicus
TITLE *formal name Aquifex aeolicus
ORGANISM *formal name Aquifex aeolicus
DATE 08-May-1998 *sequence_revision 06-May-1998 *text_change
08-May-1998
ACCESSIONS B70367
REFERENCE B70367
AUTHORS Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.;
Lennox, A.L.; Graham, D.E.; Overbeek, R.; Sneed, M.A.;
Keller, M.; AuJay, M.; Huber, R.; Feldman, R.A.; Short,
J.M.; Olson, J.J.; Swanson, R.V.
JOURNAL Nature (1998) 392:353-358
TITLE The complete genome of the hyperthermophilic bacterium
Aquifex aeolicus.
CROSS-REFERENCES M01D:98196666
ACCESSION B70367
STATUS preliminary: nucleic acid sequence not shown:
translation not shown
MOLECULE_TYPE DNA
RESIDUES 1-480 *label AQF

```



Search completed: Mon Jun 19 16:00:43 2000  
Job time : 10 secs.



\*\*\*\*\*  
[W][O][R][D][S] [T][M]  
\*\*\*\*\*

Release 3.1A John F. Collins, BioComputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPSrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Mon Jun 19 15:59:18 2000; MasPar time 3.43 seconds  
Tabular output not generated. 274,849 Million cell updates/sec

Title: >US-09-142-524A-4  
Description: (1-31) from US09142524A.pep  
Perfect Score: 225  
Sequence: 1:FSKNENIKLNMPLYIAGNRRIKRVSNV:31

Scoring table: PAM 150  
Gap 11

Searched: 83957 seqs, 30454973 residues  
Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: swiss-pro:38  
::swissprot

Statistics: Mean 34.404; Variance 55.946; scale 0.615

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Query Match | Score | Length | ID           | Description            | Prod. No. |
|------------|-------------|-------|--------|--------------|------------------------|-----------|
| 1          | 118         | 52.4  | 374    | SBP_CRYJA    | SUGI BASIC PROTEIN PRE | 2.90e-08  |
| 2          | 83          | 36.9  | 598    | YMF5_YEAST   | HYPOTHETICAL 104.7 KD  | 6.32e-02  |
| 3          | 81          | 36.0  | 127    | RM05_AGNCA   | MITOCHONDRIAL 60S RIBO | 1.34e-01  |
| 4          | 81          | 36.0  | 357    | PEL1_TORAC   | PECTATE LYASE PRECURSO | 1.34e-01  |
| 5          | 81          | 36.0  | 3140   | POLG_PPVRA   | GENOME POLYPROTEIN [CO | 1.34e-01  |
| 6          | 81          | 36.0  | 3143   | POLG_PPVD    | GENOME POLYPROTEIN [CO | 1.34e-01  |
| 7          | 78          | 34.7  | 143    | YEH7_YEAST   | VERY HYPOTHETICAL 16.3 | 4.06e-01  |
| 8          | 76          | 33.8  | 532    | YR85_GAEEL   | HYPOTHETICAL 59.8 KD P | 8.38e-01  |
| 9          | 76          | 33.8  | 559    | YAZ2_METJA   | HYPOTHETICAL PROTEIN M | 8.38e-01  |
| 10         | 75          | 33.3  | 347    | RH31_YEAST   | DNA DAMAGE TOLERANCE P | 1.20e+00  |
| 11         | 74          | 32.9  | 90     | AX68_SOTBN   | AUXIN-INDUCED PROTEIN  | 1.71e+00  |
| 12         | 74          | 32.9  | 92     | ARG7_PHAAR   | INDOLE-3-ACETIC ACID   | 1.71e+00  |
| 13         | 74          | 32.9  | 189    | Y573_METJA   | HYPOTHETICAL PROTEIN M | 1.71e+00  |
| 14         | 73          | 32.4  | 343    | RTA_RAT      | PROBABLE G PROTEIN-COU | 2.42e+00  |
| 15         | 73          | 32.4  | 3125   | POLG_PPVNA   | GENOME POLYPROTEIN [CO | 2.42e+00  |
| 16         | 72          | 32.0  | 238    | RM08_YEAST   | MITOCHONDRIAL 60S RIBO | 3.43e+00  |
| 17         | 72          | 32.0  | 309    | YIRB_SHIEL   | VIROLENCE REGULON TRAN | 3.43e+00  |
| 18         | 72          | 32.0  | 3140   | POLG_PPVSK   | GENOME POLYPROTEIN [CO | 3.43e+00  |
| 19         | 71          | 31.6  | 136    | R17A_YEAST   | 40S RIBOSOMAL PROTEIN  | 4.83e+00  |
| 20         | 71          | 31.6  | 202    | YC63_HUMAN   | HYPOTHETICAL PROTEIN K | 4.83e+00  |
| 21         | 71          | 31.6  | 255    | 1 MODE HAEIN | MOLYBDENUM TRANSPORT P | 4.83e+00  |
| 22         | 71          | 31.6  | 906    | R1RL_HCMVA   | R-BONUCLEOSIDE-DIPHOSP | 4.83e+00  |
| 23         | 70          | 31.1  | 327    | Y013_NPVAC   | HYPOTHETICAL 39.7 KDP  | 6.78e+00  |

|    |    |      |      |   |            |                        |          |
|----|----|------|------|---|------------|------------------------|----------|
| 24 | 70 | 31.1 | 746  | 1 | TAGE_BACSU | TEICHOIC ACID BIOSYNTH | 6.78e+00 |
| 25 | 70 | 31.1 | 1518 | 1 | POLG_PPVEA | GENOME POLYPROTEIN [CC | 6.78e+00 |
| 26 | 69 | 30.7 | 247  | 1 | YACE_ECOLI | HYPOTHETICAL 28.3 KD P | 9.49e+00 |
| 27 | 69 | 30.7 | 333  | 1 | IF2B_RABIT | EUKARYOTIC TRANSLATION | 9.49e+00 |
| 28 | 69 | 30.7 | 333  | 1 | IF2B_HUMAN | EUKARYOTIC TRANSLATION | 9.49e+00 |
| 29 | 69 | 30.7 | 351  | 1 | IGNA_PSES2 | INCREASED GLYPHOSATE R | 9.49e+00 |
| 30 | 69 | 30.7 | 502  | 1 | KSO_RAT    | RIBOSOMAL PROTEIN S6 K | 9.49e+00 |
| 31 | 69 | 30.7 | 502  | 1 | KSO_HUMAN  | RIBOSOMAL PROTEIN S6 K | 9.49e+00 |
| 32 | 69 | 30.7 | 686  | 1 | MX1A_SHIFL | MX1A PROTEIN (VIRH PRO | 9.49e+00 |
| 33 | 68 | 30.2 | 376  | 1 | Y501_MYCTU | HYPOTHETICAL 41.1 KD P | 1.32e+01 |
| 34 | 68 | 30.2 | 395  | 1 | Y164_METJA | HYPOTHETICAL PROTEIN M | 1.32e+01 |
| 35 | 68 | 30.2 | 899  | 1 | YM31_YEAST | HYPOTHETICAL 103.6 KD  | 1.32e+01 |
| 36 | 68 | 30.2 | 1018 | 1 | CONT_HUMAN | CONTACTIN PRECURSOR (G | 1.32e+01 |
| 37 | 68 | 30.2 | 1020 | 1 | CONT_MOUSE | CONTACTIN PRECURSOR (N | 1.32e+01 |
| 38 | 67 | 29.8 | 92   | 1 | AX13_SOYBN | AUXIN-INDUCED PROTEIN  | 1.64e+01 |
| 39 | 67 | 29.8 | 181  | 1 | RL11_MEDSA | 60S RIBOSOMAL PROTEIN  | 1.84e+01 |
| 40 | 67 | 29.8 | 184  | 1 | RL12_ARATH | 60S RIBOSOMAL PROTEIN  | 1.84e+01 |
| 41 | 67 | 29.8 | 256  | 1 | BLAB_BACCE | BETA-LACTAMASE, TYPE I | 1.84e+01 |
| 42 | 67 | 29.8 | 392  | 1 | MP14_AMBAR | POLLEN ALLERGEN AMB A  | 1.84e+01 |
| 43 | 67 | 29.8 | 427  | 1 | Y428_METJA | HYPOTHETICAL PROTEIN M | 1.84e+01 |
| 44 | 67 | 29.8 | 1039 | 1 | YLA6_CAEEL | HYPOTHETICAL 121.8 KO  | 1.84e+01 |
| 45 | 67 | 29.8 | 1092 | 1 | DPOG_DROME | DNA POLYMERASE DELTA C | 1.84e+01 |

ALIGNMENTS

RESULT 1

ID SBP\_CRYJA STANDARD: PRT: 374 AA.

AC P18632:

DI 01-NOV-1990 (Rel. 16, Created)

DI 01-NOV-1995 (Rel. 32, Last sequence update)

DI 15-DEC-1998 (Rel. 37, Last annotation update)

DE SUGI BASIC PROTEIN PRECURSOR (SBP) (MAJOR ALLERGEN CRY J 1) (CRY J 1).

OS Cryptomeria japonica (Japanese cedar)

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Coniferopsida; Coniferales; Taxodiaceae; Cryptomeria.

RC [1]

RP SEQUENCE FROM N.A. AND PAK. AT SEQUENCE.

RC TISSUE-POLLEN:

RX MEDLINE: 94183234.

RA Sone T., Komiyama N., Shimizu K., Kusakabe T., Morikubo K., Kino K.

RT Cloning and sequencing of cDNA coding for Cry j 1, a major allergen of Japanese cedar pollen.

RL Biochem. Biophys. Res. Commun. 199 619-625(1994).

RN [2]

RP SEQUENCE FROM N.A.

RC TISSUE-POLLEN:

RA Namba M., Karose M., Terioka K., Fukuda S., Kurimoto M.

RN [3]

RP Submitted (201-1994) to the EMBL/GenBank/DBS databases.

RN [3]

RP SEQUENCE OF 22-41.

RC TISSUE-POLLEN:

RX MEDLINE: 89031257.

RA Tanai M., Ando S., Usui M., Kurimoto M., Sakaguchi M., Inouye S., Matuhashi T.

RT N-terminal amino acid sequence of a major allergen of Japanese cedar pollen (Cry j 1).

RL FEBS Lett. 239:329-332(1988).

RN [4]

RP CARBOHYDRATES

RX MEDLINE: 95003748.

RA Hijikata A., Matsumoto T., Kojima K., Ogawa H.

RT Antigenicity of the oligosaccharide moiety of the Japanese cedar (Cryptomeria japonica) pollen allergen, Cry j1.

RL Int. Arch. Allergy Immunol. 105:198-202(1994).

CC -!- PTM: CONTAINS FUCOSE/XULOSE-CONTAINING N-LINKED OLIGOSACCHARIDES.

CC -!- DISEASE: THIS PROTEIN IS THE MAJOR ALLERGEN OF JAPANESE CEDAR POLLEN. THE MOST COMMON POLLEN ALLERGEN IN JAPAN.

CC -!- MISCELLANEOUS: THE SEQUENCE OF CRY J 1 FORM A IS SHOWN HERE. FORM B DIFFERS IN SIX POSITIONS.

CC -!- SIMILARITY: BELONGS TO THE POLYSACCHARIDE LYASE FAMILY 1.



CC AT THEIR NON-REDUCING ENDS.  
 CC -:- DEVELOPMENTAL STAGE: MAXIMALLY EXPRESSED LATE IN POLIOMYELITIS  
 CC -:- SIMILARITY: BELONGS TO THE POLYSACCHARIDE LYASE FAMILY 1.  
 CC  
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: X67158; CAA47630.1; .  
 CC EMBL: X67159; CAA47631.1; .  
 CC EMBL: X61102; CAA41414.1; .  
 CC PIR: S26211; S25211; .  
 CC PFAM: PF00544; pect\_lyase.1.  
 CC PRINTS: PR00807; AMBALLERJEN.  
 CC Lysase: Signal.  
 CC SIGNAL 1 25 POTENTIAL.  
 CC CHAIN 26 397 PECTATE LYASE.  
 CC ACT\_SITE 272 272 POTENTIAL.  
 CC CARBOHYD 134 134 POTENTIAL.  
 CC CARBOHYD 227 227 POTENTIAL.  
 CC CONFLICT 156 156 S -> C (IN MRNA).  
 CC CONFLICT 189 190 GS -> SG (IN MRNA).  
 CC CONFLICT 200 200 S -> D (IN MRNA).  
 CC CONFLICT 202 203 H -> R (IN MRNA).  
 CC CONFLICT 249 249 H -> N (IN MRNA).  
 CC CONFLICT 249 249 H -> N (IN MRNA).  
 CC SEQUENCE 397 AA; 44351 MW; EF0A82CE5DA7543F CRC64;  
 CC  
 CC Query Match 36.0%; Score 81; DB 1; Length 397;  
 CC Best Local Similarity 50.0%; Pred. No. 1; 1e-01;  
 CC Matches 10; Conservative 7; Mismatches 3; Indels 0; Gaps 0;  
 CC  
 CC 115 IFGKMKIKSLRELIVTSNK 135  
 CC 1 IFSKNLNKLNMPLYIAGNK 20  
 CC  
 CC RESULT 5  
 CC ID POLG\_PVPA STANDARD: PRT: 3140 AA.  
 CC AC P17767;  
 CC DT 01-FEB-1994 (Rel. 15, Created)  
 CC DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 CC DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 CC DE GENOME POLYMERASE (CONTAINS: N-TERMINAL PROTEIN (P1); HELPER  
 CC COMPONENT PROTEINASE (EC 3.4.22.-) (HC-PRO); PROTEIN P3; 6 KD PROTEIN  
 CC 2 (6K2); CYTOPLASMIC INCLUSION PROTEIN (CI); 6 KD PROTEIN 2 (6K2);  
 CC GENOME-LINKED PROTEIN (VPG); NUCLEAR INCLUSION PROTEIN A (NI-A) (NIA)  
 CC (EC 3.4.22.-) (49 KD PHOSPHATASE); (49 KD-PRO); NUCLEAR INCLUSION  
 CC PROTEIN B (NI-B) (NIB) (RNA-DIRECTED RNA POLYMERASE) (EC 2.7.7.48);  
 CC COAT PROTEIN (CP).  
 CC OS Plum pox virus (strain Rakovic) (ppv).  
 CC OC Viruses; ssRNA positive-strand viruses, no DNA stage; Potyviridae;  
 CC CC Potyvirus.  
 CC [1]  
 CC RP SEQUENCE FROM N.A.  
 CC RX MEDLINE: 89370814.  
 CC RA Lain S., Riechmann J.L., Garcia J.A.;  
 CC ST "The complete nucleotide sequence of plum pox potyvirus RNA.";  
 CC BL Virus Res. 13:157-172(1989).  
 CC EN  
 CC PP SEQUENCE OF 1778-2342 FROM N.A.  
 CC RX MEDLINE: 89268456.  
 CC RA Garcia J.A., Riechmann J.L., Lain S.;  
 CC RT "Proteolytic activity of the plum pox potyvirus NIA-like protein in  
 CC Escherichia coli.";  
 CC RL Virology 170:362-369(1989).  
 CC RN [3];  
 CC RP SEQUENCE OF 2263-3140 FROM N.A.  
 CC RA Lain S., Riechmann J.L., Mendez E., Garcia J.A.;

RT "Nucleotide sequence of the terminal region of plum pox potyvirus  
 RNA.";  
 RL Virus Res. 10:325-342(1988).  
 CC -:- FUNCTION: CYTOPLASMIC INCLUSION PROTEIN HAS HELICASE ACTIVITY 11  
 CC MAY BE INVOLVED IN REPLICATION.  
 CC -:- FUNCTION: NUCLEAR INCLUSION PROTEIN A HAS PROTEOLYTIC ACTIVITY.  
 CC -:- PTM: VPG IS COVALENTLY LINKED TO THE GENOMIC RNA.  
 CC -:- PTM: THE VIRAL RNA OF POTYVIRUSES IS EXPRESSED AS A SINGLE  
 CC POLYPROTEIN WHICH UNDERGOES POSTTRANSLATIONAL PROTEOLYTIC  
 CC PROCESSING RESULTING IN THE PRODUCTION OF AT LEAST EIGHT  
 CC INDIVIDUAL PROTEINS.  
 CC -:- SIMILARITY: HC PROTEINASES BELONGS TO PEPTIDASE FAMILY C6  
 CC -:- SIMILARITY: NI-A PROTEINASES BELONGS TO PEPTIDASE FAMILY C4.  
 CC -:- SIMILARITY: BELONGS TO THE POTYVIRUSES POLYPROTEIN FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: M21847; AA85458.1; .  
 CC EMBL: M26965; AA47085.1; .  
 CC PIR: A50009; GNVSR.  
 CC PFAM: PF00270; DEAD.1.  
 CC PFAM: PF00680; RNA\_dep\_RNA\_pol.1.  
 CC PFAM: PF00767; Poty\_coat.1.  
 CC PFAM: PF00851; Peptidase\_C6.1.  
 CC PFAM: PF00863; Peptidase\_C4.1.  
 CC PFAM: PF01577; Poty\_P1.1.  
 CC PRINTS: PR00966; NIAPOTYPIASE.  
 CC Hydrolase: Transferase; Triol protease; RNA-directed RNA polymerase;  
 CC Coat protein; Polyprotein; Covalent protein-RNA linkage; Helicase;  
 CC ATP-binding.  
 CC CHAIN 1 301 N-TERMINAL PROTEIN.  
 CC CHAIN 302 914 HELPER COMPONENT PROTEINASE.  
 CC CHAIN 915 ? PROTEIN P3.  
 CC CHAIN 1168 ? 6 KD PROTEIN 1.  
 CC CHAIN 1169 1803 CYTOPLASMIC INCLUSION PROTEIN.  
 CC CHAIN 1804 1856 6 KD PROTEIN 2.  
 CC CHAIN 1857 ? GENOME-LINKED PROTEIN.  
 CC CHAIN 2292 ? NUCLEAR INCLUSION PROTEIN A.  
 CC CHAIN 2293 2810 NUCLEAR INCLUSION PROTEIN B.  
 CC CHAIN 2811 3140 COAT PROTEIN.  
 CC BINDING 1919 1919 COVALENT LINKAGE OF VIRAL RNA (BY  
 CC SIMILARITY).  
 CC NP\_BIND 1253 1260 ATP (POTENTIAL).  
 CC SEQUENCE 3140 AA; 355577 MW; EF308B07942C3F5 CRC64.  
 CC  
 CC Query Match 36.0%; Score 81; DB 1; Length 3140;  
 CC Best Local Similarity 46.7%; Pred. No. 1; 34e-01;  
 CC Matches 14; Conservative 8; Mismatches 5; Indels 3; Gaps 3;  
 CC  
 CC 2410 IFS-ALNKAAGVGLY-SGKKRDYFKNVSD 2437  
 CC 1 IFSKNLNKLNMPLYIAGNKRRFKRVSN 29  
 CC  
 CC RESULT 6  
 CC ID POLG\_PVPA STANDARD: PRT: 3141 AA.  
 CC AC P13529; Q84929; P89038;  
 CC DT 01-JAN-1990 (Rel. 13, Created)  
 CC DT 15-FEB-2000 (Rel. 39, Last sequence update)  
 CC DT 15-FEB-2000 (Rel. 39, Last annotation update)  
 CC DE GENOME POLYMERASE (CONTAINS: N-TERMINAL PROTEIN (P1); HELPER  
 CC COMPONENT PROTEINASE (EC 3.4.22.-) (HC-PRO); PROTEIN P3; 6 KD PROTEIN  
 CC 1 (6K1); CYTOPLASMIC INCLUSION PROTEIN (CI); 6 KD PROTEIN 2 (6K2);  
 CC GENOME-LINKED PROTEIN (VPG); NUCLEAR INCLUSION PROTEIN A (NI-A) (NIA)  
 CC (EC 3.4.22.-) (49 KD PHOSPHATASE); (49 KD-PRO); NUCLEAR INCLUSION  
 CC PROTEIN B (NI-B) (NIB) (RNA-DIRECTED RNA POLYMERASE) (EC 2.7.7.48);  
 CC COAT PROTEIN (CP).  
 CC OS Plum pox virus (strain Rakovic) (ppv).  
 CC OC Viruses; ssRNA positive-strand viruses, no DNA stage; Potyviridae;  
 CC CC Potyvirus.  
 CC [1]  
 CC RP SEQUENCE FROM N.A.  
 CC RX MEDLINE: 89370814.  
 CC RA Lain S., Riechmann J.L., Garcia J.A.;  
 CC ST "The complete nucleotide sequence of plum pox potyvirus RNA.";  
 CC BL Virus Res. 13:157-172(1989).  
 CC EN  
 CC PP SEQUENCE OF 1778-2342 FROM N.A.  
 CC RX MEDLINE: 89268456.  
 CC RA Garcia J.A., Riechmann J.L., Lain S.;  
 CC RT "Proteolytic activity of the plum pox potyvirus NIA-like protein in  
 CC Escherichia coli.";  
 CC RL Virology 170:362-369(1989).  
 CC RN [3];  
 CC RP SEQUENCE OF 2263-3140 FROM N.A.  
 CC RA Lain S., Riechmann J.L., Mendez E., Garcia J.A.;

US Plum pox potyvirus (strain D) (PPV).  
 CC Viruses: ssRNA positive-strand viruses, no DNA stage; Potyviridae.  
 CC Potyvirus.

PN (1)  
 RX SEQUENCE FROM N.A.  
 RX MEDLINE: 90056790.  
 RA Teycheney P.Y., Taveri G., Delbos R., Ravelonandro M., Dunez J.,  
 RI "The complete nucleotide sequence of plum pox virus RNA (strain D)."  
 RL Nucleic Acids Res. 17:1015-1016 (1989).

PN (2)  
 RX SEQUENCE OF 2810-3141 FROM N.A. AND SEQUENCE OF 2812-2828.  
 RA Ravelonandro M., Varveris C., Delbos R., Dunez J.,  
 RI "Nucleotide sequence of the capsid protein gene of plum pox  
 RT potyvirus".  
 RL J. Gen. Virol. 69:1509-1516 (1988).

RN (13)  
 RP REVISIONS TO C-TERMINUS.

RA le Gall G.;

RP Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.

CC FUNCTION: HELPER COMPONENT-PROTEINASE IS REQUIRED FOR APHID

CC TRANSMISSION AND A.S.U. HAS PROTEOLYTIC ACTIVITY.

CC FUNCTION: CYTOPLASMIC INCLUSION PROTEIN HAS HELICASE ACTIVITY. IT

CC MAY BE INVOLVED IN REPLICATION.

CC FUNCTION: NUCLEAR INCLUSION PROTEIN A HAS PROTEOLYTIC ACTIVITY.

CC P.N. VPG IS COVALENTLY LINKED TO THE GENOMIC RNA.

CC P.N. THE VIRAL RNA OF POTYVIRUSES IS EXPRESSED AS A SINGLE

CC POLYPEPTIDE WHICH UNDERGOES POSTTRANSLATIONAL PROTEOLYTIC

CC PROCESSING RESULTING IN THE PRODUCTION OF AT LEAST EIGHT

CC INDIVIDUAL PROTEINS.

CC SIMILARITY: NC PROTEINASE BELONGS TO PEPTIDASE FAMILY C6.

CC SIMILARITY: NI-A PROTEINASE BELONGS TO PEPTIDASE FAMILY C4.

CC SIMILARITY: BELONGS TO THE POTYVIRUSES POLYPEPTIDE FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed, usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: X16415; CAA34437.1;

DR EMBL: 000299; BAB002.0.1;

DR FIC: S05929; GNVSPD.

DR FIC: S0079; S0078.

DR PFAM: PF00270; TRAC1.

DR PFAM: PF00590; RNA\_dep\_PNA\_P1.

DR PFAM: PF00767; Potyvirus\_P1.

DR PFAM: PF00951; Eupluidase\_P1.

DR PFAM: PF03563; Eupluidase\_P2.

DR PFAM: PF01577; Potyvirus\_P1.

DR PRIN2: PR00966; NACPPTASE.

KW Hydrolase; Transferase; Thiol protease; RNA-directed RNA polymerase;

KW Coat protein; Polyprotein; Covalent protein-RNA linkage; Helicase;

KW ATP-binding.

FT CHAIN 1 301 N-TERMINAL PROTEIN.

FT CHAIN 322 915 HELPER COMPONENT PROTEINASE.

FT CHAIN 316 2 6 KD PROTEIN 1.

FT CHAIN ? 1169 PROTEIN P3.

FT CHAIN 1170 1804 CYTOPLASMIC INCLUSION PROTEIN.

FT CHAIN 1205 1937 6 KD PROTEIN 2.

FT CHAIN 1858 2293 GENOME-LINKED PROTEIN.

FT CHAIN ? 2293 NUCLEAR INCLUSION PROTEIN A.

FT CHAIN 2294 2811 NUCLEAR INCLUSION PROTEIN B.

FT CHAIN 2812 3141 COAT PROTEIN.

FT CHAIN 1920 1920 COVALENT LINKAGE OF VIRAL RNA (BY

FT SIMILARITY).

FT NP\_S:ND 1254 1261 ATP (POTENTIAL).

FT SEQUENCE 3141 AA: 355569 MW: 633641C404414DBB CRC64:

Query Match 36.0% Score 81 DB 1: Length 3141:

Res: Local Similarity 46.7% Pred: NC 1.34e-01:

Matches 14: Conservative 8: Mismatches 5: Indels 3: Gaps 3:

DB 2421 IFS-ALNKAAGALY-SOKKROYEKNVSD 2438

QY 1 IFSKNLNTKLN-PLYAGNKRFF-KRVSN 29

QY 1 IFSKNLNTKLN-PLYAGNKRFF-KRVSN 29

QY 1 IFSKNLNTKLN-PLYAGNKRFF-KRVSN 29

QY 1 IFSKNLNTKLN-PLYAGNKRFF-KRVSN 29

QY 1 IFSKNLNTKLN-PLYAGNKRFF-KRVSN 29

QY 1 IFSKNLNTKLN-PLYAGNKRFF-KRVSN 29

QY 1 IFSKNLNTKLN-PLYAGNKRFF-KRVSN 29

QY 1 IFSKNLNTKLN-PLYAGNKRFF-KRVSN 29

QY 1 IFSKNLNTKLN-PLYAGNKRFF-KRVSN 29

QY 1 IFSKNLNTKLN-PLYAGNKRFF-KRVSN 29

QY 1 IFSKNLNTKLN-PLYAGNKRFF-KRVSN 29

QY 1 IFSKNLNTKLN-PLYAGNKRFF-KRVSN 29

QY 1 IFSKNLNTKLN-PLYAGNKRFF-KRVSN 29

QY 1 IFSKNLNTKLN-PLYAGNKRFF-KRVSN 29

QY 1 IFSKNLNTKLN-PLYAGNKRFF-KRVSN 29

QY 1 IFSKNLNTKLN-PLYAGNKRFF-KRVSN 29

QY 1 IFSKNLNTKLN-PLYAGNKRFF-KRVSN 29

QY 1 IFSKNLNTKLN-PLYAGNKRFF-KRVSN 29

QY 1 IFSKNLNTKLN-PLYAGNKRFF-KRVSN 29

QY 1 IFSKNLNTKLN-PLYAGNKRFF-KRVSN 29

QY 1 IFSKNLNTKLN-PLYAGNKRFF-KRVSN 29

QY 1 IFSKNLNTKLN-PLYAGNKRFF-KRVSN 29

QY 1 IFSKNLNTKLN-PLYAGNKRFF-KRVSN 29

QY 1 IFSKNLNTKLN-PLYAGNKRFF-KRVSN 29

QY 1 IFSKNLNTKLN-PLYAGNKRFF-KRVSN 29

QY 1 IFSKNLNTKLN-PLYAGNKRFF-KRVSN 29

QY 1 IFSKNLNTKLN-PLYAGNKRFF-KRVSN 29

QY 1 IFSKNLNTKLN-PLYAGNKRFF-KRVSN 29

QY 1 IFSKNLNTKLN-PLYAGNKRFF-KRVSN 29

QY 1 IFSKNLNTKLN-PLYAGNKRFF-KRVSN 29

QY 1 IFSKNLNTKLN-PLYAGNKRFF-KRVSN 29

QY 1 IFSKNLNTKLN-PLYAGNKRFF-KRVSN 29

QY 1 IFSKNLNTKLN-PLYAGNKRFF-KRVSN 29

QY 1 IFSKNLNTKLN-PLYAGNKRFF-KRVSN 29

QY 1 IFSKNLNTKLN-PLYAGNKRFF-KRVSN 29

QY 1 IFSKNLNTKLN-PLYAGNKRFF-KRVSN 29

QY 1 IFSKNLNTKLN-PLYAGNKRFF-KRVSN 29

QY 1 IFSKNLNTKLN-PLYAGNKRFF-KRVSN 29

QY 1 IFSKNLNTKLN-PLYAGNKRFF-KRVSN 29

QY 1 IFSKNLNTKLN-PLYAGNKRFF-KRVSN 29

QY 1 IFSKNLNTKLN-PLYAGNKRFF-KRVSN 29

QY 1 IFSKNLNTKLN-PLYAGNKRFF-KRVSN 29

QY 1 IFSKNLNTKLN-PLYAGNKRFF-KRVSN 29

QY 1 IFSKNLNTKLN-PLYAGNKRFF-KRVSN 29

QY 1 IFSKNLNTKLN-PLYAGNKRFF-KRVSN 29

QY 1 IFSKNLNTKLN-PLYAGNKRFF-KRVSN 29

QY 1 IFSKNLNTKLN-PLYAGNKRFF-KRVSN 29

QY 1 IFSKNLNTKLN-PLYAGNKRFF-KRVSN 29

QY 1 IFSKNLNTKLN-PLYAGNKRFF-KRVSN 29

QY 1 IFSKNLNTKLN-PLYAGNKRFF-KRVSN 29

QY 1 IFSKNLNTKLN-PLYAGNKRFF-KRVSN 29

QY 1 IFSKNLNTKLN-PLYAGNKRFF-KRVSN 29

QY 1 IFSKNLNTKLN-PLYAGNKRFF-KRVSN 29

QY 1 IFSKNLNTKLN-PLYAGNKRFF-KRVSN 29

QY 1 IFSKNLNTKLN-PLYAGNKRFF-KRVSN 29

QY 1 IFSKNLNTKLN-PLYAGNKRFF-KRVSN 29

QY 1 IFSKNLNTKLN-PLYAGNKRFF-KRVSN 29

QY 1 IFSKNLNTKLN-PLYAGNKRFF-KRVSN 29

QY 1 IFSKNLNTKLN-PLYAGNKRFF-KRVSN 29

QY 1 IFSKNLNTKLN-PLYAGNKRFF-KRVSN 29

QY 1 IFSKNLNTKLN-PLYAGNKRFF-KRVSN 29

QY 1 IFSKNLNTKLN-PLYAGNKRFF-KRVSN 29

QY 1 IFSKNLNTKLN-PLYAGNKRFF-KRVSN 29

QY 1 IFSKNLNTKLN-PLYAGNKRFF-KRVSN 29

QY 1 IFSKNLNTKLN-PLYAGNKRFF-KRVSN 29

QY 1 IFSKNLNTKLN-PLYAGNKRFF-KRVSN 29

CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DB EMBL: U23515; AAC45645.1;  
 DR WORKPEP: R144.5; CE02035;  
 KW Hypothetical protein.  
 SQ SEQUENCE 532 AA: 59809 MW: 103588FF445B4283 CRC64:

Query Match 33.8% Score 76; DB 1; Length 532;  
 Best Local Similarity 35.5% Pred. No. 8.38e-01;  
 Matches 11; Conservative 11; Mismatches 7; Indels 2; Gaps 2;

DB 160 LFSNANK-KLVSPVLC-NRPLRIRHTDIV 208  
 QY 1 LFSKNLNKLNMPYAGNKRFRKRVSVI 31

RESULT 9  
 ID YA22-METCA STANDARD: PRT: 569 AA.  
 AC Q58428;  
 DI 15-JUL-1998 (Rel. 36, Created)  
 DI 15-JUL-1998 (Rel. 36, Last sequence update)  
 DI 15-JUL-1999 (Rel. 36, Last annotation update)  
 DE HYPOTHETICAL PROTEIN M31022.  
 GN M31022.  
 OS Methanococcus jannaschii  
 CC Archaea: Euryarchaeota: Methanococcales: Methanococcaceae:  
 CC Methanococcus.  
 CC [1]

SEQUENCE FROM N.A.  
 STRAIN=JAL-1 / DSM 2661 / ATCC 43067;  
 MEDLINE: 96317599.  
 RA Bait C.J., White G., Olson G.J., Zhou L., Fleischmann R.D.,  
 RA Sitten G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Goynne J.D.,  
 RA Kariavade A.R., Dougherty B.A., Tob J.-F., Adams K.D., Reich C.L.,  
 RA Overbeck R., Kirkness E.F., Weinstein K.G., Metrick J.M., Glodek A.,  
 RA Scott J.L., Geoghagan N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,  
 RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,  
 RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,  
 RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;  
 RA "Complete genome sequence of the methanogenic archaeon, Methanococcus  
 jannaschii";  
 SC Science 273:1058-1073(1996).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DB EMBL: U67545; AAB93026.1;  
 DR TIGR: M31022;  
 DR PFAM: PF01472; PUA; 1.  
 KW Hypothetical protein.  
 SQ SEQUENCE 569 AA: 55581 MW: E29EC27908E366ED CRC64:

Query Match 33.8% Score 76; DB 1; Length 569;  
 Best Local Similarity 40.7% Pred. No. 8.38e-01;  
 Matches 11; Conservative 8; Mismatches 6; Indels 2; Gaps 2;

DB 102 FSKHVDLIEADLYIADGRLLER 128  
 QY 2 FSKNLN-IKL-NMPLYAGNKRFRKRV 26

RESULT 10  
 ID RH31-YFAST STANDARD: PRT: 347 AA.  
 AC Q06624;

DI 01-NOV-1997 (Rel. 35, Created)  
 DI 01-NOV-1997 (Rel. 35, Last sequence update)  
 DI 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE DNA DAMAGE TOLERANCE PROTEIN RHC31 (RAD31 HOMOLOG).  
 GN AOS1 OR RHC31 OR YER185W OR Y97051.5.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 CC Eukaryota: Fungi: Ascomycota: Saccharomycetes: Saccharomycetidae:  
 CC Saccharomycetaceae: Saccharomycetes.  
 CC [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=S288C / AB972;  
 RA Compton M., Andrews S., Brinkman N., Cooper J., Ding H., Du Z.,  
 RA Favetto A., Fulton L., Gattung S., Greco T., Kirsten J.,  
 RA Kucaba T., Wallisworth K., Hawkins J., Hillier L., Jier M.,  
 RA Johnson D., Johnston L., Lanston Y., Latreille P., Le T.,  
 RA Mardis E., Meneses S., Miller N., Nhan M., Pauley A., Peluso D.,  
 RA Rifkin L., Riles L., Raich A., Trevaskis E., Vignati D.,  
 RA Wilcox L., Mohdram P., Vaudin M., Wilson R., Waterston R.;  
 RL Submitted (MAY-1995; to the EMBL/GenBank/DBJ databases.  
 RN [2]

CHARACTERIZATION:  
 RX MEDLINE: 97248684.  
 RA Shayeghi M., Does C.L., Tavassoli M., Watts F.;  
 RI "Characterisation of Schizosaccharomyces pombe rad31, a USA-related  
 RI gene required for DNA damage tolerance";  
 RL Nucleic Acids Res. 25:1162-1169(1997).  
 CC -1- FUNCTION: COULD BE INVOLVED IN A UBIQUITIN-RELATED PROCESS  
 CC IMPORTANT FOR DNA DAMAGE TOLERANCE.  
 CC -1- SIMILARITY: TO THE N-TERMINAL OF UBIQUITIN-ACTIVATING ENZYME E1.  
 CC -1- SIMILARITY: STRONG, TO S-POMBE RAD31.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: U25842; AAB68113.1;  
 DR SGD: L0004031; AOS1.  
 DR PFAM: PF00899; Thif\_family; 1.  
 KW DNA damage. 347 AA: 39273 MW: 1AD87B817RA277F3E CRC64:  
 SQ SEQUENCE

Query Match 33.3% Score 76; DB 1; Length 347;  
 Best Local Similarity 50.0% Pred. No. 1.20e-02;  
 Matches 10; Conservative 6; Mismatches 5; Indels 1; Gaps 0;

DB 140 NTLARKLNIPLYVAGSNLF 159  
 QY 4 KNLNKLNPPLYAGNKRFR 23

RESULT 11  
 ID AX6B-SOYBN STANDARD: PRT: 90 AA.  
 AC P33083;  
 DI 01-OCT-1993 (Rel. 27, Created)  
 DI 01-OCT-1993 (Rel. 27, Last sequence update)  
 DI 01-OCT-1993 (Rel. 27, Last annotation update)  
 DE AUXIN-INDUCED PROTEIN 6B.  
 OS Glycine max (Soybean)  
 CC Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta;  
 CC euphyllophytes: Spermatophyta: Magnoliophyta: eudicotyledons;  
 CC core eudicots: Rosidae: eurosids 1: Fabales: Fabaceae: Papilionoideae:  
 CC Glycine.  
 CC [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV WAYNE;  
 RX MEDLINE: 92404712.  
 RA McClure B.A., Hagen G., Brown C.S., Gee M.A., Guilfoyle T.J.;  
 RI "Transcription, organization, and sequence of an auxin-regulated gene  
 RI cluster in soybean";

```

DE 01-NOV-1997 (Rel. 35, Created)
DI 01-NOV-1997 (Rel. 35, Last sequence update)
DI 01-NOV-1997 (Rel. 35, Last annotation update)
DE HYPOTHETICAL PROTEIN M03573.
GN M03573.
OC Methanococcus jannaschii.
OC Archaea: Euryarchaeota: Methanococcales: Methanococcaceae.
OC Methanococcus.
RN [1]
RP SEQUENCE FROM N.A.
RC S-RAIN-JAL-1 / JSM 2661 / ATCC 43047.
RX MEDLINE: 96337999.
RA Butt C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton K.A., Ketchum I.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.F., Adams M.D., Reith G.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodok A.,
RA Scott J.L., Geoghegan N.S.M., Weisman J.F., Ehrmann J.L., Nguyen D.,
RA Otterback T.R., Kerley J.M., Peterson J.D., Sadow P.W., Hanks M.D.,
RA Cotton M.D., Roberts K.M., Hurt M.A., Kalne B.P., Borodovsky M.,
RA Kieck H.-P., Fraser C.M., Smith R.D., Woese C.R., Venter J.C.
RA *Complete genome sequence of the methanogenic archaeon, Methanobrevan-
RI tianaschii.
RL Science 273:1056-1073(1996).
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U67506; AAB98573.1; .
CC --GR: M30573; .
CC KW Hypothetical protein.
CC SQ SEQUENCE 189 AA; 21198 MW; 4717693ALDERF69 CRC64.

Query Match 32.9% Score 74; DB 1; Length 92;
Best Local Similarity 33.3% Pred. No. 1,71e-00;
Matches 10; Conservative 11; Mismatches 7; Indels 1; Gaps 1;

DB 20 SKVDLPKGYLAVYVGEMKRFVIPVSHL 48
QY 3 SKNLNI-KLNPLYIAGNKRKFVPSNV 30

RESULT 12
ID ARG7 PHAAU STANDARD; PRI: 52 AA
AC P32295.
SI 01-OCT-1993 (Rel. 27, Created)
DI 01-OCT-1993 (Rel. 27, Last sequence update)
DI 01-OCT-1993 (Rel. 27, Last annotation update)
DE INDOLE-3-ACETIC ACID INDUCED PROTEIN ARG7.
GN ARG7.
OS Phaseolus aureus (Mung bean) (Vigna radiata).
OC Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta:
OC euphyllophytes: Spermatophyta: Magnoliophyta: eudicotyledons:
OC core eudicots: Rosidae: eufrosids I: Fabales: Fabaceae: Papilionoideae:
OC Vigna.
CC [1]
CC
CC SEQUENCE FROM N.A.
CC TISSUE-HYPOCOTYL.
CC Yamamoto K.I., Mori H., Imaseki H.
CC cDNA cloning of indole-3-acetic acid regulated ononin Aox22 and SAUR
CC from mung bean (Vigna radiata) hypocotyl tissue.
CC Plant Cell Physiol. 33:93-97(1992).
CC
CC -!- DEVELOPMENTAL STAGE: FUND IN ELONGATING HYPOCOTYLS
CC -!- INDUCED BY AUXIN AND CYCLOHEXIMIDE.
CC -!- SIMILARITY: BELONGS TO THE ARG7 FAMILY.
CC
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: D14414; BAA03310.1; .
CC SEQUENCE 92 AA; 10212 MW; 3ED5A9CEBCF699E CRC64.

Query Match 32.9% Score 74; DB 1; Length 92;
Best Local Similarity 34.5% Pred. No. 1,71e-00;
Matches 10; Conservative 11; Mismatches 7; Indels 1; Gaps 1;

DB 20 SKVDLPKGYLAVYVGEMKRFVIPVSHL 48
QY 3 SKNLNI-KLNPLYIAGNKRKFVPSNV 30

RESULT 13
ID Y573.MET-2A STANDARD; PRI: 189 AA.
AC Q37933.

```

```

DI 01-NOV-1997 (Rel. 35, Created)
DI 01-NOV-1997 (Rel. 35, Last sequence update)
DI 01-NOV-1997 (Rel. 35, Last annotation update)
DE HYPOTHETICAL PROTEIN M03573.
GN M03573.
OC Methanococcus jannaschii.
OC Archaea: Euryarchaeota: Methanococcales: Methanococcaceae.
OC Methanococcus.
RN [1]
RP SEQUENCE FROM N.A.
RC S-RAIN-JAL-1 / JSM 2661 / ATCC 43047.
RX MEDLINE: 96337999.
RA Butt C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton K.A., Ketchum I.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.F., Adams M.D., Reith G.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodok A.,
RA Scott J.L., Geoghegan N.S.M., Weisman J.F., Ehrmann J.L., Nguyen D.,
RA Otterback T.R., Kerley J.M., Peterson J.D., Sadow P.W., Hanks M.D.,
RA Cotton M.D., Roberts K.M., Hurt M.A., Kalne B.P., Borodovsky M.,
RA Kieck H.-P., Fraser C.M., Smith R.D., Woese C.R., Venter J.C.
RA *Complete genome sequence of the methanogenic archaeon, Methanobrevan-
RI tianaschii.
RL Science 273:1056-1073(1996).
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U67506; AAB98573.1; .
CC --GR: M30573; .
CC KW Hypothetical protein.
CC SQ SEQUENCE 189 AA; 21198 MW; 4717693ALDERF69 CRC64.

Query Match 32.9% Score 74; DB 1; Length 189;
Best Local Similarity 33.3% Pred. No. 1,71e-00;
Matches 11; Conservative 11; Mismatches 9; Indels 2; Gaps 2;

DB 22 MPMKMPALALEKFLOVIPENKKEFLARVQII 54
QY 1 IFSKNLNKLNWPL-VIAGNKR-F-KPVSNI 31

RESULT 14
ID KIA-PAI STANDARD; PRI: 143 AA
AC P23749.
SI 01-NOV-1993 (Rel. 26, Created)
DI 01-NOV-1993 (Rel. 26, Last sequence update)
DI 01-JUL-1993 (Rel. 26, Last annotation update)
DE PROBABLE G PROTEIN-COUPLED RECEPTOR RTA.
GN RTA.
OS Rattus norvegicus (Rat).
OC Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Mammalia:
OC Eutheria: Rodentia: Sciurognathi: Muridae: Murinae: Rattus
OC [1]
CC SEQUENCE FROM N.A.
CC STRAIN-SPRAGUE-DAWLEY; TISSUE-ACRTA.
CC MEDLINE: 90222168.
CC Ross P.C., Figler R.A., Corlay M.H., Barber C.M., Adam N.,
CC Marcus D.R., Lynch K.R.
CC *RTA, a candidate G protein-coupled receptor: cloning, sequencing,
CC and tissue distribution.
CC Proc. Natl. Acad. Sci. U.S.A. 87:3052-3056(1990).
CC
CC -!- FUNCTION: ORPHAN RECEPTOR.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -!- TISSUE SPECIFICITY: GUT, VAS DEFERENS, UTERUS, AND AGRTA BUT ONLY
CC BARELY DETECTABLE IN LIVER, KIDNEY, LUNG, AND SALIVARY GLAND. IN
CC THE BRAIN, RTA IS MARKEDLY ABUNDANT IN THE CEREBELLUM.
CC -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC MOST SIMILAR TO MAS.

```

```

CC-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement. (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

```

CC POLYPROTEIN WHICH UNDERGOES POSTTRANSLATIONAL PROTEOLYTIC
CC PROCESSING RESULTING IN THE PRODUCTION OF AT LEAST EIGHT
CC INDIVIDUAL PROTEINS.
CC
CC -1- SIMILARITY: HC PROTEINASE BELONGS TO PEPTIDASE FAMILY C6.
CC -1- SIMILARITY: N-A PROTEINASE BELONGS TO PEPTIDASE FAMILY C4
CC -1- SIMILARITY: BELONGS TO THE POLYVIRUSES POLYPROTEIN FAMILY
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by not-for-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to licens@isb-sib.ch).
CC
CC EMBL: D13751; BAA02898.1;
CC PIR: JQ0003; GNVSP.
CC
CC PFAM: PF00270; DEAD; 1.
CC PFAM: PF00680; RNA_dep_RNA_pol;
CC PFAM: PF00767; Poly_coat; 1.
CC PFAM: PF00851; Peptidase_C6; 1.
CC PFAM: PF00863; Peptidase_C4; 1.
CC PFAM: PF01577; Poly_E1; 1.
CC
CC PRINTS: PRO0966; NIAPOTYPEASE.
CC
CC Hydrolase; Transferase; Thiol protease; RNA-directed RNA polymerase;
CC Coat protein; Polyprotein; Covalent protein-RNA linkage; Helicase.
CC
CC AIP-binding.
CC
CC CHAIN 1 30; N-TERMINAL PROTEIN.
CC CHAIN 302 914; 15 PER COMPONENT PROTEINASE.
CC CHAIN 915 ?; ...TEIN P3.
CC
CC CHAIN ? 1168; 6 KD PROTEIN 1.
CC CHAIN 1169 1803; CYTOPLASMIC INCLUSION PROTEIN
CC CHAIN 1804 1856; 6 KD PROTEIN 2.
CC CHAIN 1857 ?; GENOME-LINKED PROTEIN
CC CHAIN ? 2292; NUCLEAR INCLUSION PROTEIN A.
CC CHAIN 2293 2810; NUCLEAR INCLUSION PROTEIN B.
CC CHAIN 2811 3125; COAT PROTEIN.
CC CHAIN 1919 1919; COVALENT LINKAGE OF VIRAL RNA (RY
CC BINDING; SIMILARITY).
CC
CC NP_BIND 1253 1260; AIP (POENTIAL).
CC SEQUENCE 3125 AA; 354261 MW; EDCDB3C439CB712 CRC64;
CC
CC Query Match 32.48; Score 73; DB 1; Length 3125;
CC Best Local Similarity 43.38; Pred. No. 2.42e-00;
CC Matches 13; Conservative 8; Mismatches 6; Indels 3; Gaps 3;
CC
CC Db 2410 IFS-ALNVAACVALLY-SKKKKDYFENVSQ 2417
CC
CC QY 1 IFSKNUNIKLNM-PLYIAGKKRFRKFSN 29
CC
CC Search completed: Mon Jun 19 15:59:38 2000
CC Job time : 20 secs.

```

```

Query Match      32.48; Score 73; DB 1; Length 3125;
Best Local Similarity 43.38; Pred. No. 2.42e+00;
Matches         13; Conservative 8; Mismatches 6; Indels 3; Gaps 3;

```

Db 2410 ZFS-ALNKAAGALY-SKKKROYEENVSQ 24.17  
CV 1 ZFSKNUN.KLNM-PLYIAGNKRRF.KRVSZ 29

1 : FSKNJN : KUNM - PLYIACNKKR : KRVSN 23

Search completed: Mon Jun 19 15:59:38 2000  
Job time : 20 secs.

```

Job time : 20 secs.

```







```

GN B9C24:
OC Borrelia burgdorferi (Lyme disease spirochete).
OC Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
RN [..]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 35210 / B31;
RX MEDLINE: 98065943.
RA FRASER C.M., CASSENS S., HUANG W.M., SUTTON G.G., CLAYTON R.A.,
RA LATHIGRA R., WHITE O., KETCHUM K.A., DOOSON R., HICKEY E.K., GWINN M.,
RA DOUGHERTY B., TOMB J.F., FEISCHMAN R.D., RICHARDSON D.,
RA PETERSON J., KERLAVAGE A.R., QUACKENBUSH J., SALZBERG S., HANSON M.,
RA VAN VEGT R., PALMER N., ADAMS M.D., GOCAYNE J.D., WEIDMAN J.,
RA UTTERBACK T., WATTEY L., McDONALD L., ARTIACH P., BOWMAN C.,
RA GARLAND S., FUJII C., COTTON M.D., HORST K., ROBERTS K., HAICH B.,
RA SMITH H.C., VENTER J.C.,
RI "Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi."
RI Nature 390:580-585(1997).
RL EMBL: AE001137; AAC66691.1; .
DL TIGR: BB0261; .
KW Hypothetical protein.
SQ SEQUENCE 460 AA: 54024 MW: 54143655 CRC32:
Query Match 36.0%; Score 81; DB 2; Length 460;
Best Local Similarity 33.3%; Pred. No. 4,17e-00;
Matches 10; Conservative 10; Mismatches 10; Indels 0; Gaps 0;

DE 320 FSKKNENIALRIYLRSKKEFKIANEII 349
QY 2 FSKNNINIKNMPLYIAGNKRFRKRVSNVI 31

RESULT 7
ID Q9XH58 PRELIMINARY: PRI: 740 AA.
AC Q9XH58:
DI 01-NOV-1999 (TREMBLrel. 12, Created)
DI 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DI 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
DE ETHYLENE RECEPTOR HOMOLOG.
UN PHE11.
OS Pelargonium hortorum.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Magnoliophyta; eud; Oryledons;
OC core eudicots; Rosidae; Geraniales; Geraniaceae; Pelargonium.
RN [..]
RP SEQUENCE FROM N.A.
RC DERVINIS C., CLARK D.G.;
RI "Effect of Pollination and Exogenous Ethylene on Accumulation of ETR;
RI Homolog Transcript During Flower Petal Abscission of Petunia
RI (Pelargonium x hortorum (L. H. Bailey))."
RI Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
DL EMBL: AF141928; RAD3576.1; .
KW RECEPTOR.
SQ SEQUENCE 740 AA: 82936 MW: 96130075 CRC32:
Query Match 34.7%; Score 78; DB 10; Length 740;
Best Local Similarity 30.8%; Pred. No. 1,21e-00;
Matches 8; Conservative 10; Mismatches 6; Indels 0; Gaps 0;

DE 443 LNVSPDLPEYVIGDEKRLVQIMLVV 468
QY 6 LNIKLNMPLYIAGNKRFRKRVSNVI 31

RESULT 8
ID Q46181 PRELIMINARY: PRI: 73 AA.
AC Q46181:
DI 01-JUN-1998 (TREMBLrel. 06, Created)
DI 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
DI 01-MAY-1999 (TREMBLrel. 10, Last annotation update)
DE HYPOTHETICAL 8.5 KD PROTEIN (FRAGMENT).
OS Drosophila subobscura (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

```

```

OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
RN [..]
RP SEQUENCE FROM N.A.
RC STRAIN=H211;
RA BLESA D., MARTINEZ-SEBASTIAN M.J.;
RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
DL EMBL: U73032; AAB92390.1; .
DR FLYBASE: FBgn0023240; DsubNanon1.
KW Hypothetical protein.
FT NON-TER.
SQ SEQUENCE 73 AA: 8614 MW: 1F41786E CRC32:
Query Match 34.2%; Score 77; DB 5; Length 73;
Best Local Similarity 28.6%; Pred. No. 1,72e-00;
Matches 8; Conservative 12; Mismatches 8; Indels 0; Gaps 0;

DE 38 KKLNQVDMSLFFAGALSFSVLSINSIF 65
QY 4 KNLNKLKMPLYIAGNKRFRKRVSNVI 31

RESULT 9
ID Q9XE03 PRELIMINARY: PRI: 256 AA.
AC Q9XE03:
DI 01-NOV-1999 (TREMBLrel. 12, Created)
DI 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DI 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
DE ORF2 (FRAGMENT).
OS Staphylococcus aureus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Staphylococcus.
RN [..]
RP SEQUENCE FROM N.A.
RC STRAIN=NCIC8325;
RX MEDLINE: 99279010.
RA ITO T., KATAYAMA Y., HIRAMATSU K.;
RI "Cloning and nucleotide sequence determination of the entire and DNA
RI of pre-methicillin-resistant Staphylococcus aureus N315."
RI Antimicrob. Agents Chemother. 43:1449-1458(1999).
DL EMBL: AB014440; BAA82253.1; .
FT NON-TER.
SQ SEQUENCE 256 AA: 30603 MW: F1FABCB3 CRC32:
Query Match 34.2%; Score 77; DB 2; Length 256;
Best Local Similarity 29.0%; Pred. No. 1,72e-00;
Matches 9; Conservative 10; Mismatches 10; Indels 0; Gaps 0;

DE 184 LFSKTMIGESLSIRVFENPKLPLKRIYV 214
QY 1 LFSKNENIKNMPLYIAGNKRFRKRVSNVI 31

RESULT 10
ID Q65965 PRELIMINARY: PRI: 480 AA.
AC Q65965:
DI 01-AUG-1998 (TREMBLrel. 07, Created)
DI 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
DI 01-AUG-1998 (TREMBLrel. 08, Last annotation update)
DE HYPOTHETICAL 56.0 KD PROTEIN.
GN AQ_765.
OS Aquifex aeolicus.
OC Bacteria; Aquificales; Aquificaceae; Aquifex.
RN [..]
RP SEQUENCE FROM N.A.
RC STRAIN=VF5;
RX MEDLINE: 98196666.
RA DECKERT G., WARREN P.V., GAASTERLAND T., YOUNG W.G., LENOX A.L.,
RA GRAHAM D.E., OVERBECK R., SNEAD M.A., KELLER M., AUJAY M., HUBER R.,
RA FELDMAN R.A., SHORT J.M., OLSON G.J., SWANSON R.V.;
RI "The complete genome of the hyperthermophilic bacterium Aquifex
RI aeolicus."
RI Nature 392:353-358(1998).

```



Search completed: Mon Jun 19 16:00:15 2000  
Job time : 20 secs.

```

RESULT 14
ID P97528 PRELIMINARY: PRI: 1028 AA.
AC P97528
DT 01-MAY-1997 (Tremblrel: 03, Created)
DI 01-MAY-1997 (Tremblrel: 03, Last sequence update)
DI 01-NOV-1999 (Tremblrel: 12, Last annotation update)
DE NB-3.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN 14.
RP SEQUENCE FROM N.A.
RC STRAIN-WISIAK; TISSUE=SPRAIN;
RX MEDLINE: 97101230.
RA OGAWA J., KANEKO H., MASUDA T., NAGATA S., HOSoya H., KATANABE K.:
RT "Novel neural adhesion molecules in the Contactin/F3 subgroup of the
RT immunoglobulin superfamily: isolation and characterization of cDNAs
RT from rat brain."
RL Neurosci. Lett. 218:173-176(1996).
DR EMBL: D87248; BAA13320.1; -.
DR HSSP: P20241; ICFB.
DR PFAM: PF00041; fn3; 4.
DR PFAM: PF00047; ig; 5.
SC SEQUENCE 1028 AA; 114065 MW; 4798F057 CRC32;

Query Match 33.3%; Score 75; DB 11; Length 1028;
Best Local Similarity 58.3%; Pred. No. 3.44e+00;
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

DB 163 PLYVQEDKRRFV 174
QY 11 PLYAGNKRRI 24

RESULT 15
ID P74570 PRELIMINARY: PRI: 33; AA.
AC P74570
DT 02-FEB-1997 (Tremblrel: 02, Created)
DI 01-FEB-1997 (Tremblrel: 02, Last sequence update)
DI 01-JAN-1999 (Tremblrel: 03, Last annotation update)
DE HYPOTHETICAL 37.6 KD PROTEIN.
OS Synechocystis sp. (Strain PCC 6803)
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
RN 11.
RP SEQUENCE FROM N.A.
RC STRAIN=PCC6803;
RA TABATA S.;
RL Submitted (JUN-1996) to the EMBL/GenBank/DDBJ databases.
RN 12.
RP SEQUENCE FROM N.A.
RC STRAIN=PCC6803;
RX MEDLINE: 97051201.
RA KANEKO T., SATO S., KOTANI H., TANAKA A., ASAMIZU E., NAKAMURA Y.,
RA MIYAJIMA N., HIRASAWA M., SUGIURA M., SASAMOTO S., KIMURA T.,
RA HOSOGUCHI T., MATSUO A., MURAKI A., NAKAZAKI N., NARUO K., OKUMURA S.,
RA SHIMO S., TAKEUCHI C., WADA T., WATANABE A., YANADA M., YASUDA M.,
RA TABATA S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions."
RL DNA Res. 3:109-136(1996).
DR EMBL: D90516; BAA18677.1; -.
KW Hypothetical protein.
SQ SEQUENCE 331 AA; 37593 MW; B50A9356C CRC32;

Query Match 32.9%; Score 74; DB 2; Length 331;
Best Local Similarity 32.0%; Pred. No. 4.84e+00;
Matches 8; Conservative 12; Mismatches 5; Indels 0; Gaps 0;

DB 4 NYOFELSFAGEDREYVDVANLL 28
QY 11 KLNPLIAGNKRRIKRVSNVI 31

```

\*\*\*\*\*  
WATERMANN  
\*\*\*\*\*  
(TM)

Release 3.1A John F. Collins, BioComputing Research Unit.  
Copyright: (c) 1994-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd  
March\_pp protein - protein database search, using Smith-Waterman algorithm  
Run On: Mon Jun 19 16:03:25 2000: Maspar time 4.36 seconds  
Tabular output not generated. 168,264 Million cell updates/sec

Title: XUS-09-142-524A-5  
Description: (1-31) from US09-142524A pep  
Perfect Score: 207  
Sequence: 1 SSGKNEGNIYNNNEAFKVERRFKRVSNVI 31  
Scoring table: PAM 150  
Gap 1:  
Searched: 189463 seqs, 23686106 residues  
Post-processing: Minimum Match 0%  
Listing first 45 summaries  
Database: a-geneseq35  
l-genesep  
Statistics: Mean 22.687: Variance 86.099: scale 0.264

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description                    | Pred. No. |
|------------|-------|-------------|--------|-------|--------------------------------|-----------|
| 1          | 207   | 100.0       | 31     | 1     | W27373 Multi-epitope peptide   | 1.63e-12  |
| 2          | 125   | 60.4        | 20     | 1     | W42154 T-cell epitope peptide  | 3.41e-04  |
| 3          | 125   | 60.4        | 354    | 1     | W03344 Chamaecyparis obtusa p  | 3.41e-04  |
| 4          | 125   | 60.4        | 354    | 1     | W42121 Japanese cypress polie  | 3.41e-04  |
| 5          | 125   | 60.4        | 375    | 1     | W03345 Chamaecyparis obtusa p  | 3.41e-04  |
| 6          | 138   | 52.2        | 367    | 1     | R45577 Jun s                   | 1.49e-02  |
| 7          | 93    | 44.9        | 20     | 1     | R82522 Cry j I Japanese Cedar  | 3.80e-01  |
| 8          | 93    | 44.9        | 20     | 1     | R45573 Cry j I pollen allergen | 3.80e-01  |
| 9          | 93    | 44.9        | 20     | 1     | W83058 Control peptide P16     | 3.80e-01  |
| 10         | 93    | 44.9        | 22     | 1     | W80255 Sugil allergen protein  | 3.80e-01  |
| 11         | 93    | 44.9        | 26     | 1     | R45599 Cry j I pollen allergen | 3.80e-01  |
| 12         | 93    | 44.9        | 28     | 1     | R92180 Cry j I Japanese Cedar  | 3.80e-01  |
| 13         | 93    | 44.9        | 30     | 1     | R45598 Cry j I pollen allergen | 3.80e-01  |
| 14         | 93    | 44.9        | 353    | 1     | R75362 Japanese cedar pollen   | 3.80e-01  |
| 15         | 93    | 44.9        | 353    | 1     | R81587 Cedar pollen allergen   | 3.80e-01  |
| 16         | 93    | 44.9        | 374    | 1     | R31937 Cry j I                 | 3.80e-01  |
| 17         | 93    | 44.9        | 374    | 1     | R60165 Japanese cedar pollen   | 3.80e-01  |
| 18         | 93    | 44.9        | 374    | 1     | R45541 Cry j I pollen allergen | 3.80e-01  |
| 19         | 93    | 44.9        | 374    | 1     | W82490 Cry j I Japanese Cedar  | 3.80e-01  |
| 20         | 90    | 43.5        | 81     | 1     | W80358 Sugil allergen protein  | 7.17e-01  |
| 21         | 99    | 43.0        | 19     | 1     | Y01802 Peptide used for the t  | 8.96e-01  |
| 22         | 87    | 42.0        | 19     | 1     | Y01801 Peptide used for the t  | 1.35e-00  |
| 23         | 85    | 41.1        | 18     | 1     | Y01804 Peptide used for the t  | 2.04e-00  |

|    |    |      |      |   |                                |          |
|----|----|------|------|---|--------------------------------|----------|
| 24 | 83 | 40.1 | 18   | 1 | Y01803 Peptide used for the t  | 3.09e-00 |
| 25 | 82 | 38.6 | 31   | 1 | W27372 Multi-epitope peptide   | 3.80e-00 |
| 26 | 80 | 38.6 | 24   | 1 | R45600 Cry j I pollen allergen | 5.73e-00 |
| 27 | 80 | 38.6 | 26   | 1 | R45601 Cry j I pollen allergen | 5.73e-00 |
| 28 | 76 | 36.7 | 80   | 1 | W27369 Multi-epitope peptide   | 1.29e-01 |
| 29 | 76 | 36.7 | 105  | 1 | W27370 Multi-epitope peptide   | 1.29e-01 |
| 30 | 76 | 36.7 | 134  | 1 | W27371 Multi-epitope peptide   | 1.29e-01 |
| 31 | 71 | 34.3 | 292  | 1 | R04895 Penicillinase-Inhibitor | 4.51e-01 |
| 32 | 70 | 33.8 | 360  | 1 | R45666 Thermotoga OC1/4V endo  | 4.28e-01 |
| 33 | 70 | 33.8 | 350  | 1 | W49870 Thermotoga OC1/4V endo  | 4.28e-01 |
| 34 | 69 | 33.3 | 50   | 1 | W59334 Haemophilus influenzae  | 5.20e-01 |
| 35 | 68 | 32.9 | 20   | 1 | W42155 T-cell epitope peptide  | 6.33e-01 |
| 36 | 68 | 32.9 | 264  | 1 | Y00910 S. aureus ffb protein   | 6.33e-01 |
| 37 | 68 | 32.9 | 455  | 1 | Y00910 S. aureus ffb protein   | 6.33e-01 |
| 38 | 67 | 32.4 | 14   | 1 | R91585 Cedar pollen allergen   | 7.69e-01 |
| 39 | 67 | 32.4 | 14   | 1 | Y01805 Peptide used for the t  | 7.69e-01 |
| 40 | 66 | 31.9 | 428  | 1 | W15543 50K-cellulase from Mel  | 9.33e-01 |
| 41 | 66 | 31.9 | 581  | 1 | R13490 Human C4 binding prote  | 9.33e-01 |
| 42 | 66 | 31.9 | 3457 | 1 | R62504 Large polypeptide sequ  | 9.33e-01 |
| 43 | 65 | 31.4 | 130  | 1 | R94166 Nematode excretory pro  | 1.13e-02 |
| 44 | 65 | 31.4 | 497  | 1 | W37735 Cytochrome P450RA1 iso  | 1.13e-02 |
| 45 | 65 | 31.4 | 1266 | 1 | W48721 Human cytoplasmic iso   | 1.13e-02 |

ALIGNMENTS

RESULT 1  
ID W27373 standard; peptide: 31 AA.  
AC W27373:  
DT 24-MAR-1998 (first entry)  
DE Multi-epitope peptide used as immunotherapeutic agent #5.  
KW Multi-epitope peptide; immunotherapeutic agent; allergic disease;  
KW T-cell epitope region; allergen; lymphocyte; immunoglobulin E.  
OS Synthetic.  
PN W09732600-A1.  
PD 12-SEP-1997 J00740.  
PE 10-MAR-1997 J00740.  
PR 10-MAR-1996: JP-080702.  
PA (MEIP ) MEIJI MILK PROD CO Ltd.  
PI Daifiki K, Iwama A, Kino K, Kame A, Sone T.  
DR WPI: 97-470495/43.  
PT Peptide immunotherapeutic agent to treat allergic diseases -  
PT contains multi-epitope peptide containing T cell epitope regions  
PT from different allergens  
PS Claim 9; Page 32: 58pp; Japanese.  
CC The present sequence represents a multi-epitope peptide which is used as  
CC a new immunotherapeutic agent. It comprises T cell epitope regions from 2  
CC or more different allergens (preferably linked via arginine or lysine  
CC dipeptide), where the T cell epitope regions have a positivity index  
CC greater than 100 as measured in a patient group responding to the  
CC allergen; have at least 75% reactivity with lymphocytes from patients  
CC responding to the allergen; and are not reactive with immunoglobulin E  
CC (IgE) antibodies from patients responsive to the allergen. The agent can  
CC be used to prevent and treat a wide variety of allergic diseases, e.g. by  
CC desensitisation. Side effects, e.g. those mediated by IgE, are reduced.  
SQ Sequence 31 AA:  
Query Match 100.0%; Score 207; DB 1; Length 31;  
Best Local Similarity 100.0%; Pred. No. 1.63e-12;  
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
DB 1 SSGKNEGNIYNNNEAFKVERRFKRVSNVI: 31  
QY 1 SSGKNEGNIYNNNEAFKVERRFKRVSNV: 31  
RESULT 2  
ID W42154 standard; peptide: 20 AA.  
AC W42154:  
DT 16-JUN-1998 (first entry)  
DE T-cell epitope peptide 34 from Japanese cypress pollen antigen (Chao).  
KW Japanese cypress pollen; antigen; T-cell epitope; Chao; Chao2;  
KW diagnosis; allergy; spring tree pollen disease; pollinosis.

```

OS Chamaecyparis obtusa.
PN W03747648-A1.
PD 18-DEC-1997.
PF 12-JUN-1997: J02031.
PR 14-JUN-1994: JP-153527.
PA (MEIP ) MEIJ MILK PROD CO LTD.
P: Dairiki K, Kino K.
DR WPI: 98-052242/05.
PT T-cell epitope peptide portion of Japanese cypress pollen antigens
PT Chaol and Chaol2 - used for diagnosis and treatment of spring tree
PT pollen disease
PS Claim 1: Page 32: 71pp: Japanese.
CC The present sequence represents a T-cell epitope peptide from Japanese
CC cypress pollen antigen Chaol. The present invention describes peptides
CC which correspond to the T-cell epitope sites on Japanese cypress pollen
CC antigens Chaol and Chaol2. The peptides can be used as a reagent for the
CC diagnosis of allergy to Japanese cypress pollen, and as an antigen in
CC the treatment and prevention of spring tree pollen disease in which the
CC pollinosis involves reactivity to Japanese cypress pollen.
SQ Sequence 20 AA:

Query Match 60.4% Score 125: DB 1: Length 20:
Best Local Similarity 100.0% Pred. No. 3,41e-04:
Matches 20: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

DE 1 SSGKNEGTNIYNNNEAFKVE 20
QY 1 SSGKNEGTNIYNNNEAFKVE 20

RESULT 5
ID W04344 standard: Protein: 354 AA.
AC W04344:
DT 28-NOV-1996 (first entry)
DE Chamaecyparis obtusa pollen allergen Cha o 1 (A).
KW Pollen allergen: Cha o 1: T-cell epitope, prevention; treatment;
KW pollinosis.
OS Chamaecyparis obtusa.
PN J08176192-A.
PD 09-JUL-1996.
PF 21-DEC-1994: JP-335089.
PR 21-DEC-1994: JP-335089.
PA (MEIP ) MEIJ MILK PROD CO LTD.
DR WPI: 98-368225/37.
DR N-PSDB: T38519.
PT DNA encoding chamaecyparis obtusa pollen allergen - T cell
PT epitope(s) of which are useful in development of preventative and
PT treating agent for C. obtusa pollen pollinosis
PS Claim 2: Pages 10-11: 17pp: Japanese.
CC The present sequence is the C. obtusa pollen allergen Cha o 1.
CC the T-cell epitopes of which can be used in the development of a
CC preventive and treating agent for C. obtusa pollen pollinosis.
CC C. obtusa pollen (2.4 kb) was decreased with diethyl ether, and
CC dried at room temp. overnight. Cha o 1 was sepd. from it and
CC purified. RNA was extracted from C. obtusa pollen, and mRNA and
CC cDNA derived.
SQ Sequence 354 AA:

Query Match 60.4% Score 125: DB 1: Length 354:
Best Local Similarity 100.0% Pred. No. 3,41e-04:
Matches 20: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

Db 311 SSGKNEGTNIYNNNEAFKVE 352
QY 1 SSGKNEGTNIYNNNEAFKVE 20

RESULT 4
ID W04344 standard: Protein: 354 AA.
AC W04344:
DT 28-JUN-1994 (first entry)
DE Japanese cypress pollen antigen Chaol.
KW Japanese cypress pollen; antigen; T-cell epitope; Chaol; Chaol2.

```

diagnosis: allergy: spring tree pollen disease; pollinosis.

OS Chamaecyparis obtusa.

PN W03747648-A1.

PD 18-DEC-1997.

PF 12-JUN-1997: J02031.

PR 14-JUN-1994: JP-153527.

PA (MEIP ) MEIJ MILK PROD CO LTD.

P: Dairiki K, Kino K.

DR WPI: 98-052242/05.

PT T-cell epitope peptide portion of Japanese cypress pollen antigens

PT Chaol and Chaol2 - used for diagnosis and treatment of spring tree

PT pollen disease

PS Example 1: Page 14-15: 7pp: Japanese.

CC The present sequence represents Japanese cypress pollen antigen Chaol.

CC The present invention describes peptides which correspond to the T-cell

CC epitope sites on Japanese cypress pollen antigens Chaol and Chaol2. The

CC peptides can be used as a reagent for the diagnosis of allergy to

CC Japanese cypress pollen, and as an antigen in the treatment and

CC prevention of spring tree pollen disease in which the pollinosis

CC involves reactivity to Japanese cypress pollen.

SQ Sequence 354 AA:

Query Match 60.4% Score 125: DB 1: Length 354:

Best Local Similarity 100.0% Pred. No. 3,41e-04:

Matches 20: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

Db 311 SSGKNEGTNIYNNNEAFKVE 350

QY 1 SSGKNEGTNIYNNNEAFKVE 20

RESULT 5

ID W04345 standard: Protein: 375 AA

AC W04345:

DT 28-NOV-1996 (first entry)

DE Chamaecyparis obtusa pollen allergen Cha o 1 CDNA (B).

KW Pollen allergen: Cha o 1: T-cell epitope; prevention; treatment;

KW pollinosis.

OS Chamaecyparis obtusa.

PN J08176192 A.

PD 09-JUL-1996.

PF 21-DEC-1994: 335089.

PR 21-DEC-1994: JP-335089.

PA (MEIP ) MEIJ MILK PROD CO LTD.

DR WPI: 96-368225/37.

DR N-PSDB: T38519.

PT DNA encoding chamaecyparis obtusa pollen allergen - T cell

PT epitope(s) of which are useful in development of preventative and

PT treating agent for C. obtusa pollen pollinosis

PS Claim 8: Pages 11-12: 17pp: Japanese.

CC The present sequence is the C. obtusa pollen allergen Cha o 1.

CC the T-cell epitopes of which can be used in the development of a

CC preventive and treating agent for C. obtusa pollen pollinosis.

CC C. obtusa pollen (2.4 kb) was decreased with diethyl ether, and

CC dried at room temp. overnight. Cha o 1 was sepd. from it and

CC purified. RNA was extracted from C. obtusa pollen, and mRNA and

CC cDNA derived.

SQ Sequence 375 AA:

Query Match 60.4% Score 125: DB 1: Length 375:

Best Local Similarity 100.0% Pred. No. 3,41e-04:

Matches 20: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

Db 332 SSGKNEGTNIYNNNEAFKVE 352

QY 1 SSGKNEGTNIYNNNEAFKVE 20

RESULT 5

ID R45577 standard: Protein: 367 AA.

AC R45577:

DT 13-JUL-1994 (first entry)

DE Japanese cypress pollen antigen Chaol.

KW - Japanese cedar: pollen allergen; allergy: treatment; diagnosis;  
KW - cell epitope; sensitivity: detection;  
OS - Juniperus sabinoides.  
EN Key Location/Qualifiers  
FI peptide 1: 21: signal peptide  
FI peptide 22: 367  
FI peptide /note= "mature peptide"  
PN W09401560-A.  
PD 20-JAN-1994.  
PF 15-JAN-1993: U00139.  
PR 10-JUL-1992: W0-J05661.  
PP 01-SEP-1992: US-938390.  
PA (IMMUNO) IMMULOGIC PHARM CORP.  
PI Bond JF, Garman RD, Griffith LJ, Kuo M, Pollock J.  
DK WPI: 94-035066/04.  
DR N-PS2B: 055272.  
PI Antigens derived from Japanese cedar pollen allergen Cry j 1.  
PI contain at least two T cell epitope(s), used to treat or diagnose  
PI allergy.  
PS Disclosure: Fig 16: 137pp; English.  
CC The sequence is that of Jun S I, a homologue of the Japanese  
CC cedar pollen allergen Cry j 1. Antigenic peptides derived from it  
CC can be used for the treatment and diagnosis of allergies associated  
CC with Japanese cedar pollen.  
SQ Sequence 367 AA:

Query Match 52.2% Score 108: DB 1: Length 367;  
Best Local Similarity 85.0% Pred. No. 1.49e-02;  
Matches 17: Conservative 1: Mismatches 2: Indels 0: Gaps 0:

DB 332 SSGKTEETNINYNNEAFKVE 351  
||||| ||||| ||||| |||||  
QY 1 SSGKNEGNIYNNNEAFKVE 20

RESULT 7  
ID R82522 standard: Protein: 20 AA.  
AC R82522.  
DE 15-APR-1996 (first entry)  
DE Cry j 1 Japanese Cedar pollen allergen peptide fragment (CJ1-32).  
KW Cry j 1: Japanese cedar pollen allergen; modified: drug production;  
KW allergy: Cryptomeria japonica.  
OS Cryptomeria japonica.  
PN W09527786-A1.  
PL 19-OCT-1995.  
PF 06-APR-1995: U04249.  
PR 08-APR-1994: US-226248.  
PK 06-SEP-1994: US-152226.  
PA (IMMUNO) IMMULOGIC PHARM CORP.  
PI Chen X, Evans S, Francis RM, Kuo M, Powers SF.  
PI Shaked Z.  
DR WPI: 95-36631/47.  
PR Modified Cryptomeria japonica (Cry j) 1 peptide(s) - useful for  
PI treating allergy to Japanese cedar pollen allergen or  
PI immunologically cross reactive allergens  
PS Disclosure: Figure 2: 60pp; English.  
CC Novel peptides of cry j 1 have been modified as a part of a  
CC preformulation scheme to develop an optimised drug product for  
CC therapeutic treatment of humans suffering from allergy to Japanese  
CC cedar pollen allergen or an allergen which is immunologically cross  
CC reactive with Japanese cedar pollen allergen. Such modified peptides  
CC possess certain characteristics which render them particularly  
CC suitable for drug product formulation. Peptide fragments of Cry j 1,  
CC modified and unmodified, are given in R82491-R82525. This peptide  
CC fragment corresponds to amino acids 311-330 of the allergen mature  
CC protein.  
SQ Sequence 20 AA:

Query Match 44.9% Score 93: DB 1: Length 20;  
Best Local Similarity 70.0% Pred. No. 3.80e-01;  
Matches 14: Conservative 3: Mismatches 3: Indels 0: Gaps 0:

DB 1 SSGKYEGGNIYTKKEAFNVE  
||||| ||||| ||||| |||||  
QY 1 SSGKNEGNIYNNNEAFKVE 20

RESULT 8  
ID R45573 standard: Protein: 20 AA.  
AC R45573.  
DE 13-JUL-1994 (first entry)  
DE Cry j 1 pollen allergen peptide CJ1-32.  
KW Japanese cedar: detection: allergy: treatment; diagnosis;  
KW T cell epitope: sensitivity.  
OS Cryptomeria japonica.  
PN W09401560-A.  
PD 20-JAN-1994.  
PF 15-JAN-1993: U00139.  
PR 10-JUL-1992: W0-C05661.  
PP 01-SEP-1992: US-938990.  
PA (IMMUNO) IMMULOGIC PHARM CORP.  
PI Bond JF, Garman RD, Griffith LJ, Kuo M, Pollock J.  
DK WPI: 94-035066/04.  
DR Antigens derived from Japanese cedar pollen allergen Cry j 1.  
DR contain at least two T cell epitope(s), used to treat or diagnose  
PI allergy.  
PS Claim 1: Fig 13: 137pp; English.  
CC The sequence is that of an isolated peptide of the Japanese cedar  
CC pollen allergen Cry j 1 (amino acids 311-330). The peptide, CJ1-32,  
CC can be used for the treatment and diagnosis of allergies associated  
CC with Japanese cedar pollen. It has enhanced therapeutic properties  
CC but reduced side effects compared to naturally occurring allergens.  
SQ Sequence 20 AA:

Query Match 44.9% Score 93: DB 1: Length 20;  
Best Local Similarity 70.0% Pred. No. 3.80e-01;  
Matches 14: Conservative 3: Mismatches 3: Indels 0: Gaps 0:

DB 1 SSGKYEGGNIYTKKEAFNVE 20  
||||| ||||| ||||| |||||  
QY 1 SSGKNEGNIYNNNEAFKVE 20

RESULT 9  
ID R83056 standard: Peptide: 20 AA.  
AC R83056.  
DE 15-MAR-1999 (first entry)  
DE Control peptide P16.  
KW Fast epitope: HFE7A; monoclonal antibody: humanised antibody;  
KW human: apoptosis: HFE7A autoimmune disease: Hashimoto's disease;  
KW systemic lupus erythematosus: viral versus host disease;  
KW Sjogren syndrome: pernicious anaemia: Addison's disease;  
KW scleroderma: Goodpasture syndrome: Crohn's disease: sterility;  
KW rheumatoid arthritis: autoimmune haemolytic anaemia;  
KW myasthenia gravis: multiple sclerosis: Basedow's disease;  
KW thrombopenia purpura: insulin-dependent diabetes: allergy;  
KW atopy: arteriosclerosis: myocarditis: cardiomyopathy: AIDS;  
KW glomerular nephritis: hypoplastic anaemia: hepatitis: AIDS;  
KW transplant rejection: therapy.  
OS Synthetic.  
PN AU9859701-A.  
PD 08-OCT-1998.  
PF 30-MAR-1998: 059701.  
PR 08-OCT-1997: JP-276064.  
PR 01-APR-1997: JP-082953.  
PR 25-JUN-1997: JP-169088.  
PA (SANYO) SANKYO CO LTD.  
PI Akio S, Hideyuki H, Hiroko Y, Jun O, Kimihisa I,  
PI Masahiko O, Nobufusa S, Shin Y, Tohru T.  
DR WPI: 98-543440/47.  
PT New antibodies and proteins bind conserved epitope of Ras antigen -  
PT used to evaluate drugs in animal models and to treat Ras-associated  
PT diseases e.g. autoimmune disease, allergy, atopy, arteriosclerosis,  
PT myocarditis, hepatitis and AIDS  
PS Reference Example 6; Page 87; 292pp; English.

CC Synthetic peptides PL-P15 (see W83043-57) are partial sequences of  
 CC amino acids 1-157 of the extracellular domain of human Fas, with  
 CC between 9 and 11 amino acid residues overlapping one another. P16  
 CC (see W83058) is a negative control having no homology with human  
 CC Fas. P1-P15 were used in an ELISA, which demonstrated that novel  
 CC murine anti-human Fas monoclonal antibody HFE7A specifically binds  
 CC an amino acid sequence contained in P11. The epitope (see W83030)  
 CC was subsequently identified. The invention provides humanised  
 CC HFE7A antibodies (see W83041-57) produced by CDR grafting. These  
 CC antibodies are capable of inducing apoptosis in abnormal cells  
 CC expressing Fas, and of inhibiting Fas-induced apoptosis in normal  
 CC cells. They are used to evaluate, in animal models, treatments of  
 CC diseases that involve Fas/Fas ligand interactions, and also to treat  
 CC such diseases.  
 CC Sequence 26 AA:

Query Match 44.9% Score 93; DB 1; Length 20;  
 Best Local Similarity 70.0%; Pred. No. 3,80e-01;  
 Matches 14; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Db 1 SSGKYECCGNIYTKKEAFNVE 20  
 QY 1 SSGKNEGTNIYNNNEAFKVE 20

## RESULT 10

ID W80355 standard; peptide; 21 AA.

AC W80355

DE 11-JAN-1999 (first entry)

CC Sugi allergen protein Cry22 derived epitope for T cells.

KW T cell epitope; sugi allergen proteins Cry11; Cry2; treatment:

KW sugi-pollinosis; allergic reaction; pollen.

OS Synthetic.

PN J1025918-A.

PD 22-SEP-1998.

PE 22-DEC-1997; 153448.

PR 24-DEC-1996; JP-34341.

PA (HAYE) HAYASHIBARA SEIBUITSU KASAKU.

FA (SANY) SANKYO CO. LTD.

DR WPI: 98-577037/49.

PT A linked T cell epitope peptide - used for the treatment of

PI Sugi-pollinosis.

PS Claim 8; pages 19-20; 21pp; Japanese.

CC W80354-58 represent epitopes for T cells, derived from the sugi allergen

CC proteins Cry11 (W80339-44; W80350-53 and W80355-58) and Cry2 (W80345-49

CC and W80354-55). The peptides are useful for the treatment of

CC sugi-pollinosis, an allergic reaction of the body to pollen.

CC Sequence 21 AA:

Query Match 44.9% Score 93; DB 1; Length 21;  
 Best Local Similarity 70.0%; Pred. No. 3,80e-01;  
 Matches 14; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Db 2 SSGKYECCGNIYTKKEAFNVE 21

QY 1 SSGKNEGTNIYNNNEAFKVE 20

## RESULT 11

ID R45595 standard; Protein; 26 AA.

AC R45595

DE 13-JUL-1994 (first entry)

CC Cry J I pollen allergen peptide CJI-44.

KW Japanese cedar; detection; allergy; treatment; diagnosis;

KW T cell epitope; sensitivity.

OS Cryptomeria japonica.

PN W09401560-A.

PD 20-JAN-1994.

PE 15-JAN-1993; U00139.

PR 10-JUL-1992; WO-003661.

FA (IMM-) IMMLOGIC PHARM CORP.

DR Bond JF, Garman RD, Griffith IJ, Kuo M, Pollock J.

DR WPI: 94-035055/04.  
 PT Antigens derived from Japanese cedar pollen allergen Cry J I  
 PT contain at least two T cell epitope(s), used to treat or diagnose  
 PT allergy.  
 PS Claim 76; Fig 18; 137pp; English.  
 CC The sequence is that of an isolated peptide of the Japanese cedar  
 CC pollen allergen Cry J I. The peptide, CJI-44.1, can be used for  
 CC the treatment and diagnosis of allergies associated with Japanese  
 CC cedar pollen. It has enhanced therapeutic properties but reduced  
 CC side effects compared to naturally occurring allergens.  
 CC Sequence 26 AA:

Query Match 44.9% Score 93; DB 1; Length 26;  
 Best Local Similarity 70.0%; Pred. No. 3,80e-01;  
 Matches 14; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Db 7 SSGKYECCGNIYTKKEAFNVE 26

QY 1 SSGKNEGTNIYNNNEAFKVE 20

## RESULT 12

ID R92180 standard; Protein; 28 AA.

AC R92180

DE 16-APR-1996 (first entry)

CC Cry J I Japanese Cedar pollen allergen modified peptide (CJI-44.2)

KW Cry J I; Japanese cedar pollen allergen; modified; drug production;

KW allergy; Cryptomeria japonica.

OS Synthetic.

PN W09527786-A.

PD 19-OCT-1995.

PE 06-APR-1995; U04249.

PR 08-APR-1994; US-226248.

PP 06-DEC-1994; US-350225.

PA (IMMU-) IMMLOGIC PHARM CORP.

PI Chen X, Evans S, Franzen HM, Kuo M, Powers SP.

PT Shaked Z.

PS Claim 2; 95-366392/47.

CC Modified Cryptomeria japonica (C J) I peptide(s) - useful for

CC treating allergy to Japanese cedar pollen allergen or

CC immunologically cross reactive allergens

PS Claim 2; Figure 3; 60pp; Eng.

CC Novel peptides of cry J I have been modified as a part of a

CC preformulation scheme to have an optimized drug product for

CC therapeutic treatment of humans suffering from allergy to Japanese

CC cedar pollen allergen or an allergen which is immunologically cross

CC reactive with Japanese cedar pollen allergen. Such modified peptides

CC possess certain characteristics which render them particularly

CC suitable for drug product formulation. In the treatment of

CC modified and unmodified are shown in Figure 1 and 2.

CC is one of three modified peptides (see Figure 1).

CC Sequence 28 AA:

Query Match 44.9% Score 94; DB 1; Length 28;  
 Best Local Similarity 70.0%; Pred. No. 3,80e-01;  
 Matches 14; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Db 9 SSGKYECCGNIYTKKEAFNVE 28

QY 1 SSGKNEGTNIYNNNEAFKVE 20

## RESULT 13

ID R45598 standard; Protein; 30 AA.

AC R45598

DE 13-JUL-1994 (first entry)

CC Cry J I pollen allergen peptide CJI-44.

KW Japanese cedar; detection; allergy; treatment; diagnosis;

KW T cell epitope; sensitivity.

OS Cryptomeria japonica.

PN W09401560-A.

PD 20-JAN-1994.

PE 15-JAN-1993; U00139.



```

PK 10-JUL-1992: WO-J05561.
PP 01-SEP-1992: US-938990.
PA (IMMUNO) IMMUNOLOGIC PHARM CORP
P1 Bond JP, Garman RD, Griffith LJ, Kuo M, Pollock J.
CP WPI: 94-035565/04.
PT Antikens derived from Japanese cedar pollen allergen Cry j 1
PI contain at least two T cell epitope(s), used to treat or diagnose
PI allergy.
PS Claim 76: Fig 18: 137pp: English.
CC The sequence is that of an isolated peptide of the Japanese cedar
CC pollen allergen Cry j 1. The peptide, CJI-44, can be used for
CC the treatment and diagnosis of allergies associated with Japanese
CC cedar pollen. It has enhanced therapeutic properties but reduced
CC side effects compared to naturally occurring allergens.
CC Sequence 36 AA:
SU
Query Match 44.9% Score 93: DB 1: Length 30:
Best Local Similarity 70.0%: Pred. No. 3.80e-01:
Matches 14: Conservative 3: Mismatches 3: Indels 0: Gaps 0:

DB 11 SSKYEGGNIYTKKAFNVE 30
QY 1 SSKKNEGNIYNNNAFKVE 20

RESULT
IL R75388 standard: protein: 353 AA.
AC R75388:1995 (first entry):
DE Japanese cedar pollen allergen Cry j 1.
KW Japanese cedar pollen allergen Cry j 1: T-cell epitope: peptides:
KW prevention: treatment: cryptomeria pollinosis.
CS Cryptomeria japonica.
SE Key Location/Qualifiers
FT peptide 61..175
F1 /note= "T-cell epitope peptide"
F2 peptide 91..105
F3 /note= "T-cell epitope peptide"
F4 peptide 106..120
F5 /note= "T-cell epitope peptide"
F6 peptide 146..160
F7 /note= "T-cell epitope peptide"
F8 peptide 211..225
F9 /note= "T-cell epitope peptide"
F10 peptide 326..340
F11 /note= "T-cell epitope peptide"
F12 peptide 355..445
F13 /note= "T-cell epitope peptide"
F14 C07116295-A.
F15 09-MAY-1995:
F16 20-OCT-1993: 254426.
F17 20-OCT-1993: JP-262126.
PA (MEIJ) MEIJY MALK 8900 CO. LTD.
DB WPI: 95-201834/27.
PI New cryptomeria pollen allergen T-cell epitope peptide - used for
PI prevention, treatment and investigation of Japanese cedar pollinosis
PS Disclosed: Figs 1-2: 8pp: Japanese.
CC R75388 is the Japanese cedar pollen allergen Cry j 1, from which the
CC T-cell epitope peptides R89289-R89295 were derived. The peptides
CC can be used for the prevention and treatment of cryptomeria
CC pollinosis, and also for the investigation of pollinosis.
CC Sequence 353 AA:
SU
Query Match 44.9% Score 93: DB 1: Length 353:
Best Local Similarity 70.0%: Pred. No. 3.80e-01:
Matches 14: Conservative 3: Mismatches 3: Indels 0: Gaps 0:

DE 311 SSKYEGGNIYTKKAFNVE 330
QY 1 SSKKNEGNIYNNNAFKVE 20

RESULT

```

```

ID R81587 standard: Protein: 353 AA.
AC R81587:1995 (first entry):
DE Cedar pollen allergen B.
KW Cedar pollen allergen: immunoglobulin E; IgE: T-cell epitope:
KW antibody: pollinosis: therapy: immunotherapy.
QS Cryptomeria japonica.
PN EP-700929-A2.
PD 13-MAR-1996.
PF 08-SEP-1995: 306295.
PR 10-SEP-1994: JP-242137.
PR 14-JUL-1995: JP-200221.
PR 14-JUL-1995: JP-200224.
PA (HAYB) HAYASHIBARA SEIBUTSU KAGAKU.
PI Hino K, Saito S, Taniguchi Y.
DR WPI: 96-140915/15.
PT New peptide(s) derived from cedar pollen allergens - activate
PT allergen-specific T-cells, but not allergen-specific IgE antibodies.
PT used for treating cedar pollinosis.
PS Claim 5: Page 31-32: 36pp: English.
CC Synthetic peptides based on portions of cedar pollen allergens A
CC (R81588) and B (R81587) were tested for their ability to activate
CC cedar allergen-specific T-cells, but not allergen-specific IgE
CC antibodies. 6 peptides (R81589-R81595) were identified as T-cell
CC epitopes. These peptides, plus subsequent peptides (R81593-79) essential
CC for T-cell recognition, and homologous peptides (R81588-96) can
CC be used as immunotherapeutic agents to treat or prevent cedar
CC pollinosis, avoiding side-effects such as anaphylaxis.
CC Sequence 353 AA:
SU
Query Match 44.9% Score 93: DB 1: Length 353:
Best Local Similarity 70.0%: Pred. No. 3.80e-01:
Matches 14: Conservative 3: Mismatches 3: Indels 0: Gaps 0:

DB 311 SSKYEGGNIYTKKAFNVE 330
QY 1 SSKKNEGNIYNNNAFKVE 20

```

Search completed: Mon Jun 19 16:03:42 2000  
 CDS time : 7 secs.

\*\*\*\*\*  
 WIRELESS (TM)  
 \*\*\*\*\*

Release 3.1A John F. Collins, BioComputing Research Unit,  
 Copyright (C) 1993-1998 University of Edinburgh, U.K.  
 Distribution Rights by Oxford Molecular Ltd

Msrch\_PP protein - protein database search, using Smith-Waterman algorithm  
 Run on: Tue Jun 20 13:34:24 2000: MasPar time 2.63 Seconds  
 Tabular output not generated. 170.246 Million cell updates/sec

Title: >US-09-142-524A-5  
 Description: (1-31) from US09142524A.pep  
 Perfect Score: 207  
 Sequence: 1 SSGKNEGTNIYNNNEAFKVERRFKRVSNVI 31

Scoring table: PAM 150  
 Gap 1:

Searched: 145341 seqs, 14437480 residues

Post-processing: Minimum Match 0%  
 Listing first 45 summaries

Database: a-issued  
 1:5A\_COMB 2:5B\_COMB 3:6\_COMB 4:PCT\_COMB 5:backfiles1

Statistics: Mean 21.454; Variance 83.581; scale 0.257

Pred. No. is the number of results predicted by chance to have a  
 score greater than or equal to the score of the suit being printed,  
 and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Query Match | Length | ID   | Description        | Pred. No.                             |
|------------|-------------|--------|------|--------------------|---------------------------------------|
| 1          | 72          | 34.8   | 417  | 2 US-08-518-615A-2 | Sequence 2, Application 30, Applicat  |
| 2          | 72          | 34.8   | 417  | 3 US-08-518-615A-2 | Sequence 115, Applicat                |
| 3          | 72          | 34.8   | 417  | 2 US-08-518-615A-2 | Sequence 115, Applicat                |
| 4          | 71          | 34.3   | 255  | 1 US-08-481-481-0  | Sequence 2, Application 30, Applicat  |
| 5          | 71          | 34.3   | 255  | 1 US-08-481-481-0  | Sequence 2, Application 30, Applicat  |
| 6          | 66          | 31.9   | 86   | 5 US-08-459-910-0  | Sequence 4, Application 2, Applicatio |
| 7          | 65          | 31.4   | 130  | 2 US-08-459-910-0  | Sequence 4, Application 2, Applicatio |
| 8          | 65          | 31.4   | 1266 | 1 US-08-459-910-0  | Sequence 4, Application 2, Applicatio |
| 9          | 64          | 30.9   | 3457 | 2 US-08-459-910-0  | Sequence 4, Application 2, Applicatio |
| 10         | 62          | 30.0   | 1045 | 1 US-08-459-910-0  | Sequence 4, Application 2, Applicatio |
| 11         | 61          | 29.5   | 411  | 1 US-08-459-910-0  | Sequence 4, Application 2, Applicatio |
| 12         | 61          | 29.5   | 500  | 1 US-08-459-910-0  | Sequence 4, Application 2, Applicatio |
| 13         | 61          | 29.5   | 538  | 1 US-08-459-910-0  | Sequence 4, Application 2, Applicatio |
| 14         | 61          | 29.5   | 541  | 1 US-08-459-910-0  | Sequence 4, Application 2, Applicatio |
| 15         | 60          | 29.0   | 128  | 3 US-08-459-910-0  | Sequence 4, Application 2, Applicatio |
| 16         | 60          | 29.0   | 128  | 3 US-08-459-910-0  | Sequence 4, Application 2, Applicatio |
| 17         | 60          | 29.0   | 248  | 1 US-08-459-910-0  | Sequence 4, Application 2, Applicatio |
| 18         | 60          | 29.0   | 248  | 1 US-08-459-910-0  | Sequence 4, Application 2, Applicatio |
| 19         | 60          | 29.0   | 249  | 2 US-08-459-910-0  | Sequence 4, Application 2, Applicatio |
| 20         | 60          | 29.0   | 249  | 2 US-08-459-910-0  | Sequence 4, Application 2, Applicatio |
| 21         | 60          | 29.0   | 249  | 2 US-08-459-910-0  | Sequence 4, Application 2, Applicatio |
| 22         | 60          | 29.0   | 249  | 2 US-08-459-910-0  | Sequence 4, Application 2, Applicatio |
| 23         | 60          | 29.0   | 261  | 1 US-08-459-910-0  | Sequence 4, Application 2, Applicatio |

| ID | US-08-518-615A-2 | STANDARD | PRT  | 317 AA            |
|----|------------------|----------|------|-------------------|
| 24 | 60               | 29.0     | 261  | 2 US-08-761-761-0 |
| 25 | 60               | 29.0     | 266  | 3 US-08-481-481-0 |
| 26 | 60               | 29.0     | 266  | 3 US-08-481-481-0 |
| 27 | 60               | 29.0     | 599  | 2 US-08-910-910-0 |
| 28 | 60               | 29.0     | 761  | 4 PCT-US93-0      |
| 29 | 60               | 29.0     | 761  | 1 US-07-906-906-0 |
| 30 | 60               | 29.0     | 761  | 1 US-08-710-710-0 |
| 31 | 60               | 29.0     | 761  | 1 US-08-192-192-0 |
| 32 | 60               | 29.0     | 1196 | 2 US-08-735-735-0 |
| 33 | 60               | 29.0     | 1196 | 2 US-08-144-144-0 |
| 34 | 59               | 28.5     | 70   | 5 524097-1        |
| 35 | 59               | 28.5     | 770  | 1 US-08-525-525-0 |
| 36 | 59               | 28.5     | 771  | 1 US-08-525-525-0 |
| 37 | 59               | 28.5     | 885  | 3 US-09-074-074-0 |
| 38 | 59               | 28.5     | 1089 | 2 US-06-742-742-0 |
| 39 | 59               | 28.5     | 1088 | 2 US-08-742-742-0 |
| 40 | 58               | 28.0     | 420  | 1 US-07-700-700-0 |
| 41 | 58               | 28.0     | 420  | 4 PCT-US92-0      |
| 42 | 58               | 28.0     | 946  | 3 US-09-074-074-0 |
| 43 | 58               | 28.0     | 3665 | 2 US-06-222-222-0 |
| 44 | 58               | 28.0     | 3712 | 2 US-08-222-222-0 |
| 45 | 58               | 28.0     | 3712 | 2 US-08-222-222-0 |

Sequence 2, Application US/08518615A  
 Patent No. 5962258  
 GENERAL INFORMATION:  
 APPLICANT: Matbur, E. et al.  
 TITLE OF INVENTION: Carboxymethyl Cellulose from Thermotoga Maritima  
 NUMBER OF SEQUENCES: 4  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,  
 ADDRESSEE: CECCHI, STEWART & OLSTEIN  
 STREET: 6 BECKER FARM ROAD  
 CITY: ROSELAND  
 STATE: NEW JERSEY  
 COUNTRY: USA  
 ZIP: 07068  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: 3.5 INCH DISKETTE  
 COMPUTER: IBM PS/2  
 OPERATING SYSTEM: MS-DOS  
 SOFTWARE: WORD PERFECT 5.1  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/518.615A  
 FILING DATE: August 23, 1995  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER:  
 FILING DATE:  
 ATTORNEY/AGENT INFORMATION:  
 NAME: FERRARO, GREGORY D.  
 REGISTRATION NUMBER: 36,134  
 REFERENCE/DOCKET NUMBER: 331400-20  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 201-994-1700  
 TELEFAX: 201-994-1744  
 INFORMATION FOR SEQ ID NO: 2:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 317 AMINO ACIDS  
 TYPE: AMINO ACID  
 STRANDEDNESS:



CC APPLICANT: RASMUSSEN, Beth A  
CC APPLICANT: TALLY, Francis P  
CC APPLICANT: GLUZMAN, Yakov  
CC TITLE OF INVENTION: DNA SEQUENCES AND AMINO ACID SEQUENCES  
CC TITLE OF INVENTION: OF CLASS B BETA-LACTAMASE ENZYMES FROM BACTERIOIDES  
CC TITLE OF INVENTION: FRAGILIS  
CC NUMBER OF SEQUENCES: 9  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Darby & Darby PC  
CC STREET: 805 Third Avenue  
CC CITY: New York  
CC STATE: NY  
CC COUNTRY: US  
CC ZIP: 10022  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: Patent Release #1.0, Version #1.30  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/459,263  
CC FILING DATE: 02-JUN-1995  
CC CLASSIFICATION: 435  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Robinson, Joseph R  
CC REGISTRATION NUMBER: 33,448  
CC REFERENCE/DOCKET NUMBER: 0646/18026-US2  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: 212-527-7700  
CC TELEFAX: 212-753-6237  
CC TELEX: 236687  
CC INFORMATION FOR SEQ ID NO: 4:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 255 amino acids  
CC TYPE: amino acid  
CC STRANDEDNESS: single  
CC MOLECULE TYPE: peptide  
CC ORIGINAL SOURCE:  
CC ORGANISM: Bacillus cereus  
CC IMMEDIATE SOURCE:  
CC CLONE: B-LACTAMASE  
CC SEQUENCE 255 AA: 27847 MW: 349604 CN:  
  
Query Match 34.38; Score 71; DB 1; Length 255;  
Best Local Similarity 57.18; Pct. No. 2.06e-01;  
Matches 8; Conservative 5; Mismatches 1; Indels 0; Gaps 0;  
  
DE 98 MVEKKFKRVTDVI 111  
QY 18 KVERRFKRVSNV 31  
  
RESULT 5  
ID US-08-459-264-4 STANDARD: PRT: 255 AA.  
XX  
AC xxxxxx  
XX  
DT  
DE Sequence 4, Application US/08459264  
XX Sequence 4, Application US/08459264  
XX Patent No. 5705340  
CC GENERAL INFORMATION:  
CC APPLICANT: RASMUSSEN, Beth A  
CC APPLICANT: TALLY, Francis P  
CC APPLICANT: GLUZMAN, Yakov  
CC TITLE OF INVENTION: DNA SEQUENCES AND AMINO ACID SEQUENCES  
CC TITLE OF INVENTION: OF CLASS B BETA-LACTAMASE ENZYMES FROM BACTERIOIDES  
CC NUMBER OF SEQUENCES: 9  
CC CORRESPONDENCE ADDRESS:

CC ADDRESSEE: Darby & Darby PC  
CC STREET: 805 Third Avenue  
CC CITY: New York  
CC STATE: NY  
CC COUNTRY: US  
CC ZIP: 10022  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: Patent Release #1.0, Version #1.30  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/459,264  
CC FILING DATE: 02-JUN-1995  
CC CLASSIFICATION: 435  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Robinson, Joseph R  
CC REGISTRATION NUMBER: 33,448  
CC REFERENCE/DOCKET NUMBER: 0546/18026-US3  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: 212-527-7700  
CC TELEFAX: 212-753-6237  
CC TELEX: 236687  
CC INFORMATION FOR SEQ ID NO: 4:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 255 amino acids  
CC TYPE: amino acid  
CC STRANDEDNESS: single  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: peptide  
CC ORIGINAL SOURCE:  
CC ORGANISM: Bacillus cereus  
CC IMMEDIATE SOURCE:  
CC CLONE: B-LACTAMASE  
CC SEQUENCE 255 AA: 27847 MW: 349604 CN:  
  
Query Match 34.38; Score 71; DB 1; Length 255;  
Best Local Similarity 57.18; Pct. No. 2.06e-01;  
Matches 8; Conservative 5; Mismatches 1; Indels 0; Gaps 0;  
  
DE 98 MVEKKFKRVTDVI 111  
QY 18 KVERRFKRVSNV 31  
  
RESULT 6  
ID 5514582-41 STANDARD: PRT: 255 AA.  
XX  
AC xxxxxx  
XX  
DT  
XX Patent No. 5514582  
XX Patent No. 5514582  
CC APPLICANT: CAPON, DANIEL J.; LASKY, LAURENCE A.  
CC TITLE OF INVENTION: RECOMBINANT DNA ENCODING HYBRID  
CC IMMUNOGENICITY:  
CC NUMBER OF SEQUENCES: 43  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/185,670  
CC FILING DATE: 21-JAN-1994  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: 986,931  
CC FILING DATE: 08-DEC-1992  
CC APPLICATION NUMBER: 603,122  
CC FILING DATE: 16-DEC-1991  
CC APPLICATION NUMBER: 440,625  
CC FILING DATE: 22-NOV-1989  
CC APPLICATION NUMBER: 315,015  
CC FILING DATE: 23-FEB-1989  
CC SEQ ID NO: 41:  
CC LENGTH: 86







CC TELEFAX: (416) 595-1163  
CC INFORMATION FOR SEQ ID NO: 18:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 538 amino acids  
CC TYPE: amino acid  
CC STRANDEDNESS: single  
CC TOPOLOGY: linear  
CC SEQUENCE 538 AA: 1483760 CN:

Query Match 29.5% Score 61: DB 2: Length 538:  
Best Local Similarity 50.0% Pred. No. 1.42e+02:  
Matches 8: Conservative 5: Mismatches 3: Indels 0: Gaps 0:

Db 6 KINPA-KVKHREIKQV 21

CC 12 NNNEAFKVERFIKRV 27

RESULT 14  
ID US-08-867-941-14 STANDARD: PRT: 541 AA.

XX xxxxxx

CC Sequence 14, Application US/08867941  
CC Patent No. 5977337

CC GENERAL INFORMATION:  
CC APPLICANT: Loosmore, Sheena M

CC APPLICANT: Du, Rui-Pan

CC APPLICANT: Wang, Qijun

CC APPLICANT: Yang, Yao-Ping

CC APPLICANT: Klein, Michael

CC TITLE OF INVENTION: LACTOFERRIN RECEPTOR GENES OF MORAXELLA

CC NUMBER OF SEQUENCES: 67

CC CORRESPONDENCE ADDRESS:

CC STREET: 6th Floor, 330 University Avenue

CC CITY: Toronto

CC STATE: Ontario

CC COUNTRY: Canada

CC ZIP: M5S 1A7

CC COMPUTER READABLE FORM:

CC MEDIUM TYPE: Floppy disk

CC COMPUTER: IBM PC compatible

CC OPERATING SYSTEM: PC-DOS/MS-DOS

CC SOFTWARE: Patent Release # 1.0, Version #1.00

CC CURRENT APPLICATION DATA:

CC APPLICATION NUMBER: US/08/867,941

CC FILING DATE: 03-JUN-1997

CC CLASSIFICATION: 435

CC ATTORNEY/AGENT INFORMATION:

CC NAME: Stewart, Michael I

CC REGISTRATION NUMBER: 24,973

CC REFERENCE/DOCKET NUMBER: 1038-681 M/S:jb

CC TELECOMMUNICATION INFORMATION:

CC TELEPHONE: (416) 595-1155

CC TELEFAX: (416) 595-1153

CC INFORMATION FOR SEQ ID NO: 14:

CC SEQUENCE CHARACTERISTICS:

CC LENGTH: 541 amino acids

CC TYPE: amino acid

CC STRANDEDNESS: single

CC TOPOLOGY: linear

CC SEQUENCE 541 AA: 50740 MW: 1494681 CN:

Query Match 29.5% Score 61: DB 2: Length 541:  
Best Local Similarity 50.0% Pred. No. 1.42e+02:  
Matches 8: Conservative 5: Mismatches 3: Indels 0: Gaps 0:

Db 6 KINPA-KVKHREIKQV 21

QY 12 NNNEAFKVERFIKRV 27

RESULT 15

ID US-08-804-439A-2 STANDARD: PRT: 128 AA.

XX xxxxxx

XX

XX

XX

DE Sequence 2, Application US/08804439A

XX Sequence 2, Application US/08804439A

CC Patent No. 6015555

CC GENERAL INFORMATION:

CC APPLICANT: Rose, Timothy M.

CC APPLICANT: Bosch, Martin L.

CC APPLICANT: Strad, Karl

CC TITLE OF INVENTION: GLYCOPROTEIN B OF THE FHV/MSHV

CC NUMBER OF SEQUENCES: 113

CC CORRESPONDENCE ADDRESS:

CC STREET: 4225 Executive Square, Ste 1400

CC CITY: La Jolla

CC STATE: CA

CC COUNTRY: USA

CC ZIP: 92037

CC COMPUTER READABLE FORM:

CC MEDIUM TYPE: Floppy disk

CC COMPUTER: IBM PC compatible

CC OPERATING SYSTEM: PC-DOS/MS-DOS

CC SOFTWARE: Patent Release #1.0, Version #1.30

CC CURRENT APPLICATION DATA:

CC APPLICATION NUMBER: US/08/804,439A

CC FILING DATE: Feb/ua 21, 1997

CC CLASSIFICATION: 424

CC ATTORNEY/AGENT INFORMATION:

CC NAME: Haller, Lisa A.

CC REGISTRATION NUMBER: 38,347

CC REFERENCE/DOCKET NUMBER: 09176/004001

CC TELECOMMUNICATION INFORMATION:

CC TELEPHONE: (619) 678-5070

CC TELEFAX: (619) 678-5069

CC TELETYPE:

CC INFORMATION FOR SEQ ID NO: 2:

CC SEQUENCE CHARACTERISTICS:

CC LENGTH: 128 amino acids

CC TYPE: amino acid

CC STRANDEDNESS: single

CC TOPOLOGY: linear

CC MOLECULE TYPE: Protein

CC SEQUENCE 128 AA: 14821 MW: 97091 CN:

Query Match 29.0% Score 60: DB 3: Length 128:

Best Local Similarity 57.1% Pred. No. 1.71e+02:

Matches 8: Conservative 3: Mismatches 2: Indels 1: Gaps 1:

Db 11 FKV-RRYIKIATSV 23

QY 17 FKVERFIKRVSNV 30

Search completed: Tue Jun 20 13:34:29 2000  
Job time: 5 secs.



\*\*\*\*\*  
WATERBURY  
\*\*\*\*\*  
(TW)

Release 3.1A John F. Collins, BioComputing Research Unit.  
Copyright (C) 1993-1998 University of Edinburgh, U.K.  
Distribution Rights by Oxford Molecular Ltd

MPsrch\_app protein - protein database search, using Smith-Waterman algorithm  
Run on: Mon Jun 19 16:23:02 2000 MasPar time 16.08 Seconds  
Tabular output not generated. 194.302 Million cell updates/sec

Title: >US-09-142-524A-5  
Description: (1-31) from US09.42524A.pep  
Sequence: 1 SSGKNEGNIYNNNEAFKVERRFKRVSNVI 31

Scoring table: FAM 150  
Gap 11

Searched: 721208 seqs, 100765575 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: a-pending  
1:PCI 2:J6 3:U60 4:U7 5:U90 6:U81 7:U82 8:U83 9:U84  
10:U84R 11:U85 12:U86 13:U87 14:U88 15:U89 16:U90 17:U91  
18:U92 19:U93 20:U94 21:U95 22:NEW0 23:NEW060 24:NEW08  
25:NEW09

Statistics: Mean 25.118; Variance 80.992; scale 0.310

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID            | Description            | Pred. No. |
|------------|-------|-------------|--------|---------------|------------------------|-----------|
| 1          | 207   | 100.0       | 31     | 17 US-09-142- | Sequence 5, Applicatio | 5.98e-01  |
| 2          | 108   | 92.2        | 346    | 3 US-65-109-  | Sequence 1, Applicatio | 1.77e-02  |
| 3          | 108   | 92.2        | 367    | 10 US-08-226- | Sequence 95, Applicati | 1.77e-02  |
| 4          | 108   | 92.2        | 367    | 10 US-08-467- | Sequence 95, Applicati | 1.77e-02  |
| 5          | 108   | 92.2        | 367    | 10 US-08-467- | Sequence 95, Applicati | 1.77e-02  |
| 6          | 108   | 92.2        | 367    | 10 US-08-350- | Sequence 95, Applicati | 1.77e-02  |
| 7          | 108   | 92.2        | 367    | 8 US-08-468-  | Sequence 95, Applicati | 1.77e-02  |
| 8          | 108   | 92.2        | 367    | 10 US-08-468- | Sequence 95, Applicati | 1.77e-02  |
| 9          | 97    | 45.9        | 30     | 10 US-08-467- | Sequence 261, Applicat | 2.60e-01  |
| 10         | 97    | 45.9        | 30     | 8 US-08-350-  | Sequence 261, Applicat | 2.60e-01  |
| 11         | 97    | 45.9        | 30     | 10 US-08-467- | Sequence 261, Applicat | 2.60e-01  |
| 12         | 97    | 45.9        | 30     | 10 US-08-468- | Sequence 261, Applicat | 2.60e-01  |
| 13         | 97    | 45.9        | 30     | 10 US-08-468- | Sequence 261, Applicat | 2.60e-01  |
| 14         | 93    | 44.9        | 20     | 4 US-07-938-  | Sequence 57, Applicati | 6.84e-01  |
| 15         | 93    | 44.9        | 20     | 8 US-08-350-  | Sequence 57, Applicati | 6.84e-01  |
| 16         | 93    | 44.9        | 26     | 7 US-08-226-  | Sequence 91, Applicati | 5.84e-01  |
| 17         | 93    | 44.9        | 26     | 8 US-08-350-  | Sequence 91, Applicati | 5.84e-01  |
| 18         | 93    | 44.9        | 25     | 10 US-08-467- | Sequence 91, Applicati | 5.84e-01  |
| 19         | 93    | 44.9        | 25     | 10 US-08-468- | Sequence 91, Applicati | 5.84e-01  |
| 20         | 93    | 44.9        | 28     | 10 US-08-468- | Sequence 132, Applicat | 5.84e-01  |

|    |    |      |     |               |                        |          |
|----|----|------|-----|---------------|------------------------|----------|
| 21 | 93 | 44.9 | 28  | 10 US-08-467- | Sequence 132, Applicat | 5.84e-01 |
| 22 | 93 | 44.9 | 28  | 10 US-08-467- | Sequence 131, Applicat | 5.84e-01 |
| 23 | 93 | 44.9 | 29  | 10 US-08-226- | Sequence 131, Applicat | 5.84e-01 |
| 24 | 93 | 44.9 | 29  | 10 US-08-467- | Sequence 131, Applicat | 5.84e-01 |
| 25 | 93 | 44.9 | 30  | 8 US-08-350-  | Sequence 260, Applicat | 5.84e-01 |
| 26 | 93 | 44.9 | 30  | 10 US-08-467- | Sequence 90, Applicati | 5.84e-01 |
| 27 | 93 | 44.9 | 30  | 10 US-08-467- | Sequence 90, Applicati | 5.84e-01 |
| 28 | 93 | 44.9 | 30  | 10 US-08-467- | Sequence 90, Applicati | 5.84e-01 |
| 29 | 93 | 44.9 | 30  | 10 US-08-467- | Sequence 90, Applicati | 5.84e-01 |
| 30 | 93 | 44.9 | 30  | 10 US-08-467- | Sequence 90, Applicati | 5.84e-01 |
| 31 | 93 | 44.9 | 30  | 10 US-08-468- | Sequence 90, Applicati | 5.84e-01 |
| 32 | 93 | 44.9 | 30  | 10 US-08-468- | Sequence 90, Applicati | 5.84e-01 |
| 33 | 93 | 44.9 | 40  | 8 US-18-350-  | Sequence 70, Applicati | 5.84e-01 |
| 34 | 93 | 44.9 | 40  | 20 US-18-467- | Sequence 70, Applicati | 5.84e-01 |
| 35 | 93 | 44.9 | 40  | 20 US-18-467- | Sequence 70, Applicati | 5.84e-01 |
| 36 | 93 | 44.9 | 40  | 70 US-226-    | Sequence 70, Applicati | 5.84e-01 |
| 37 | 93 | 44.9 | 40  | 10 US-467-    | Sequence 55, Applicati | 5.84e-01 |
| 38 | 93 | 44.9 | 63  | 10 US-350-    | Sequence 65, Applicati | 5.84e-01 |
| 39 | 93 | 44.9 | 63  | 10 US-468-    | Sequence 65, Applicati | 5.84e-01 |
| 40 | 93 | 44.9 | 63  | 10 US-08-467- | Sequence 65, Applicati | 5.84e-01 |
| 41 | 93 | 44.9 | 353 | 21 US-38-526- | Sequence 15, Applicati | 5.84e-01 |
| 42 | 93 | 44.9 | 374 | 10 US-38-468- | Sequence 2, Applicatio | 5.84e-01 |
| 43 | 93 | 44.9 | 374 | 10 US-38-468- | Sequence 2, Applicatio | 5.84e-01 |
| 44 | 93 | 44.9 | 374 | 4 US-07-729-  | Sequence 2, Applicatio | 5.84e-01 |
| 45 | 93 | 44.9 | 374 | 4 US-07-729-  | Sequence 2, Applicatio | 5.84e-01 |

ALIGNMENTS

RESULT 1  
ID US-09-142-524-5 STANDARD: PR: 41 AA.

AC xxxxxx

DT

XX

Sequence 5, Application: US/09142524

Sequence 5, Application: US/09142524

GENERAL INFORMATION:

APPLICANT: Kume, Akino

APPLICANT: Kairiki, Kazuo

APPLICANT: Kinoshita, Akio

APPLICANT: Kinoshita, Akio

TITLE OF INVENTION: Peptide based immunotherapeutic Agent For Treating

TITLE OF INVENTION: Peptide based immunotherapeutic Agent For Treating

FILE REFERENCE: Docet, No. 1997-01-14

CURRENT APPLICATION NUMBER: US/09142524

EARLIER FILING DATE: 1997-01-14

EARLIER FILING DATE: 1997-03-10

EARLIER FILING DATE: 1997-03-10

NUMBER OF SEQ ID NOS: 5

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 5

LENGTH: 31

TYPE: PRI

ORGANISM: Unknown

FEATURE:

OTHER INFORMATION: Description of Unknown Organism: peptide

Query Match 100.0% Score 207, DB 17, Length 31;  
Best Local Similarity 100.0% Fed. No. 5.98e-14;  
Matches 31; Conservative ; Mismatches 0; Indels 0; Gaps 0;

Db 1 SSGKNEGNIYNNNEAFKVERRFKRVSNVI 31

1 SSGKNEGNIYNNNEAFKVERRFKRVSNVI 31



CC FILING DATE: December 6, 1994  
 CC ATTORNEY/AGENT INFORMATION:  
 CC NAME: Jane E. Remillard  
 CC REGISTRATION NUMBER: 38,872  
 CC REFERENCE/DOCKET NUMBER: 025.6 US02 (IMI-128CPD2)  
 CC TELEPHONE: (617) 227-5941  
 CC TELEFAX: (617) 227-5941  
 CC INFORMATION FOR SEQ ID NO: 95:  
 CC SEQUENCE CHARACTERISTICS:  
 CC LENGTH: 367 amino acids  
 CC TYPE: amino acid  
 CC TOPOLOGY: linear  
 CC MOLECULE TYPE: protein  
 CC SEQUENCE 367 AA: 35846 MW: 695078 CN:  
 SQ  
 Query Match 52.2% Score 108 PB 10 Length 367;  
 Best Local Similarity 85.0% Pred. No. 1.73e-02;  
 Matches 17: Conservative 1: Mismatches 2: Indels 0: Gaps 0:  
 DB 332 SSKKTEETNIYNSNEAFKVE 351  
 IIII IIII IIII IIII  
 QY 1 SSKKNEGNIYNNNEAFKVE 20  
 RESULT 5  
 ID US-08-467-006-95 STANDARD: PRI: 367 AA.  
 XX  
 AC xxxxxx  
 DE  
 DE Sequence 95, Application US/08467006  
 XX  
 XX Sequence 95, Application: US/08467005  
 CC GENERAL INFORMATION:  
 CC APPLICANT: Griffith, Irwin J.  
 CC APPLICANT: Pollock, Joanne  
 CC APPLICANT: Bond, Julian P.  
 CC APPLICANT: Garman, Richard D.  
 CC APPLICANT: Kuo, Mei-Chang  
 CC APPLICANT: Yeung, Siu-mei H.  
 CC APPLICANT: Brauer, Andrew  
 CC APPLICANT: Exley, Mark A.  
 CC APPLICANT: Powers, Steven P.  
 CC TITLE OF INVENTION: Allergic Proteins And Peptides From  
 CC TITLE OF INVENTION: Japanese Cedar Pollen  
 CC NUMBER OF SEQUENCES: 26  
 CC CORRESPONDENCE ADDRESS:  
 CC ADDRESSEE: Immunologic Pharmaceutical Corporation, Inc.  
 CC STREET: 610 Lincoln St  
 CC CITY: Waltham  
 CC STATE: MA  
 CC COUNTRY: USA  
 CC ZIP: 02154  
 CC MEDIUM TYPE: Floppy disk  
 CC COMPUTER: IBM PC compatible  
 CC OPERATING SYSTEM: PC-DOS/MS-DOS  
 CC SOFTWARE: Patent Release #1.0, Version #1.25  
 CC CURRENT APPLICATION DATA:  
 CC APPLICATION NUMBER: US/08/457,006  
 CC FILING DATE: June 6, 1995  
 CC CLASSIFICATION: 424  
 CC PRIOR APPLICATION DATA:  
 CC APPLICATION NUMBER: 08/350,225  
 CC FILING DATE: December 6, 1994  
 CC ATTORNEY/AGENT INFORMATION:  
 CC NAME: Jane E. Remillard  
 CC REGISTRATION NUMBER: 38,872  
 CC REFERENCE/DOCKET NUMBER: 025.6 US06 (IMI-128CPD6)  
 CC TELEPHONE: (617) 227-7400

CC TELEFAX: (617) 227-5941  
 CC INFORMATION FOR SEQ ID NO: 95:  
 CC SEQUENCE CHARACTERISTICS:  
 CC LENGTH: 367 amino acids  
 CC TYPE: amino acid  
 CC TOPOLOGY: linear  
 CC MOLECULE TYPE: protein  
 CC SEQUENCE 367 AA: 35846 MW: 695078 CN:  
 SQ  
 Query Match 52.2% Score 108 PB 10 Length 367;  
 Best Local Similarity 85.0% Pred. No. 1.73e-02;  
 Matches 17: Conservative 1: Mismatches 2: Indels 0: Gaps 0:  
 DB 332 SSKKTEETNIYNSNEAFKVE 351  
 IIII IIII IIII IIII  
 QY 1 SSKKNEGNIYNNNEAFKVE 20  
 RESULT 6  
 ID US-08-467-697-95 STANDARD: PRI: 367 AA.  
 XX  
 AC xxxxxx  
 DE  
 DE Sequence 95, Application US/08467697  
 XX  
 XX Sequence 95, Application US/08467697  
 CC GENERAL INFORMATION:  
 CC APPLICANT: Griffith, Irwin J.  
 CC APPLICANT: Pollock, Joanne  
 CC APPLICANT: Bond, Julian P.  
 CC APPLICANT: Garman, Richard D.  
 CC APPLICANT: Kuo, Mei-Chang  
 CC APPLICANT: Yeung, Siu-mei H.  
 CC APPLICANT: Brauer, Andrew  
 CC APPLICANT: Exley, Mark A.  
 CC APPLICANT: Powers, Steven P.  
 CC TITLE OF INVENTION: Allergic Proteins And Peptides From  
 CC TITLE OF INVENTION: Japanese Cedar Pollen  
 CC NUMBER OF SEQUENCES: 26  
 CC CORRESPONDENCE ADDRESS:  
 CC ADDRESSEE: Immunologic Pharmaceutical Corporation, Inc.  
 CC STREET: 610 Lincoln St  
 CC CITY: Waltham  
 CC STATE: MA  
 CC COUNTRY: USA  
 CC ZIP: 02154  
 CC MEDIUM TYPE: Floppy disk  
 CC COMPUTER: IBM PC compatible  
 CC OPERATING SYSTEM: PC-DOS/MS-DOS  
 CC SOFTWARE: Patent Release #1.0, Version #1.25  
 CC CURRENT APPLICATION DATA:  
 CC APPLICATION NUMBER: US/08/467,597  
 CC FILING DATE: June 6, 1995  
 CC CLASSIFICATION: 424  
 CC PRIOR APPLICATION DATA:  
 CC APPLICATION NUMBER: 08/350,225  
 CC FILING DATE: December 6, 1994  
 CC ATTORNEY/AGENT INFORMATION:  
 CC NAME: Jane E. Remillard  
 CC REGISTRATION NUMBER: 38,872  
 CC REFERENCE/DOCKET NUMBER: 025.6 US04 (IMI-028CPD4)  
 CC TELEPHONE: (617) 227-7400  
 CC TELEFAX: (617) 227-5941  
 CC INFORMATION FOR SEQ ID NO: 95:  
 CC SEQUENCE CHARACTERISTICS:  
 CC LENGTH: 367 amino acids  
 CC TYPE: amino acid  
 CC TOPOLOGY: linear  
 CC MOLECULE TYPE: protein

SC SEQUENCE 367 AA: 39846 MW: 695078 CN:

Query Match 52.2% Score 108: DB 10: Length 367:  
 Best Local Similarity 85.0% Pred. No. 1.73e-02:  
 Matches 17: Conservative 1: Mismatches 2: Indels 0: Gaps 0:

Db 332 SSGKTEETNIYNSNEAFKVE 351  
 111 111 111 111 111  
 Qy 1 SSGKNECTNIYNNNEAFKVE 20

RESULT 7  
 ID US-09-350-225-95 STANDARD: PRI: 367 AA

XX AC xxxxxx

Sequence 95, Application US/08350225

Sequence 95, Application US/08350225

GENERAL INFORMATION:

APPLICANT: Griffith, Irwin J.

APPLICANT: Pollock, Joanne

APPLICANT: Bond, Julian F.

APPLICANT: Garman, Richard D.

APPLICANT: Kuo, Wei-Chang

APPLICANT: Yeung, Siu-Mei H.

APPLICANT: Brauer, Andrew

APPLICANT: Exley, Mark A.

APPLICANT: Powers, Steven P.

TITLE OF INVENTION: Allergenic Proteins And Peptides From

TITLE OF INVENTION: Japanese Cedar Pollen

NUMBER OF SEQUENCES: 261

CORRESPONDENCE ADDRESS:

ADDRESSEE: Immunologic Pharmaceutical Corporation, Inc.

STREET: 610 Lincoln St

CITY: Waltham

STATE: MA

COUNTRY: USA

ZIP: 02154

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #10, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/350,225

FILING DATE: December 8, 1994

CLASSIFICATION: 42

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 05/226,218

FILING DATE: April 8, 1994

APPLICATION NUMBER: 07/948,930

FILING DATE: September 1, 1992

APPLICATION NUMBER: PCI/US93/00,39

FILING DATE: January 15, 1993

ATTORNEY/AGENT INFORMATION:

NAME: Darlene A. Vanstone

REGISTRATION NUMBER: 35,729

REFERENCE/DOCKET NUMBER: 025.6 US (IMI-028CP2)

TELECOMMUNICATION INFORMATION:

TELEPHONE: (517) 466-6000

TELEFAX: (517) 466-6049

INFORMATION FOR SEQ ID NO: 95:

SEQUENCE CHARACTERISTICS:

LENGTH: 367 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE 367 AA: 39846 MW: 695078 CN:

Query Match 52.2% Score 108: DB 6: Length 367:

Best Local Similarity 85.0% Pred. No. 1.73e-02:  
 Matches 17: Conservative 1: Mismatches 2: Indels 0: Gaps 0:

Db 332 SSGKTEETNIYNSNEAFKVE 351  
 111 111 111 111 111  
 Qy 1 SSGKNECTNIYNNNEAFKVE 20

RESULT 8  
 ID US-08-468-940-95 STANDARD: PRI: 367 AA

XX AC xxxxxx

Sequence 95, Application US/08468940

Sequence 95, Application US/08468940

GENERAL INFORMATION:

APPLICANT: Griffith, Irwin J.

APPLICANT: Pollock, Joanne

APPLICANT: Bond, Julian F.

APPLICANT: Garman, Richard D.

APPLICANT: Kuo, Wei-Chang

APPLICANT: Yeung, Siu-Mei H.

APPLICANT: Brauer, Andrew

APPLICANT: Exley, Mark A.

APPLICANT: Powers, Steven P.

TITLE OF INVENTION: Allergenic Proteins And Peptides From

TITLE OF INVENTION: Japanese Cedar Pollen

NUMBER OF SEQUENCES: 261

CORRESPONDENCE ADDRESS:

ADDRESSEE: Immunologic Pharmaceutical Corporation, Inc.

STREET: 610 Lincoln St

CITY: Waltham

STATE: MA

COUNTRY: USA

ZIP: 02154

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #10, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/468,940

FILING DATE:

CLASSIFICATION: 42

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 07/948,930

FILING DATE: September 1, 1992

APPLICATION NUMBER: PCI/US93/00,39

FILING DATE: January 15, 1993

ATTORNEY/AGENT INFORMATION:

NAME: Darlene A. Vanstone

REGISTRATION NUMBER: 35,729

REFERENCE/DOCKET NUMBER: 025.6 US (IMI-028CP2)

TELECOMMUNICATION INFORMATION:

TELEPHONE: (517) 466-6000

TELEFAX: (517) 466-6049

INFORMATION FOR SEQ ID NO: 95:

SEQUENCE CHARACTERISTICS:

LENGTH: 367 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE 367 AA: 39846 MW: 695078 CN:

Query Match 52.2% Score 108: DB 10: Length 367:  
 Best Local Similarity 85.0% Pred. No. 1.73e-02:

Matches 17: Conservative 1: Mismatches 2: Indels 0: Gaps 0:

DE 332 SSGKTEETNIYNSNEAFKVE 351  
 QY 1 SSGKNEGNIYNNNEAFKVE 20

RESULT 9  
 ID US-09-467-023-261 STANDARD: PRT: 30 AA.

XX xxxxxx

DE Sequence 261, Application US/08467023

CC Sequence 261, Application US/08467023

CC GENERAL INFORMATION:

CC APPLICANT: Griffith, Irwin J.

CC APPLICANT: Pollock, Joanne;

CC APPLICANT: Bond, Julian F.

CC APPLICANT: Garman, Richard D.

CC APPLICANT: Kuo, Mei-Chang;

CC APPLICANT: Yeung, Siu-mei H.

CC APPLICANT: Brauer, Andrew;

CC APPLICANT: Exley, Mark A.

CC APPLICANT: Powers, Steven P.

CC TITLE OF INVENTION: Allergenic Proteins And Peptides From

CC TITLE OF INVENTION: Japanese Cedar Pollen

CC NUMBER OF SEQUENCES: 261

CC CORRESPONDENCE ADDRESS:

CC ADDRESSEE: Immunologic Pharmaceutical Corporation, Inc.

CC STREET: 610 Lincoln St

CC CITY: Waltham

CC STATE: MA

CC COUNTRY: USA

CC ZIP: 02154

CC COMPUTER READABLE FORM:

CC MEDIUM TYPE: Floppy disk

CC COMPUTER: IBM PC compatible

CC OPERATING SYSTEM: PC-DOS/MS-DOS

CC SOFTWARE: Patent in Release #1.0, Version #1.25

CC CURRENT APPLICATION DATA:

CC APPLICATION NUMBER: US/08467023

CC FILING DATE: June 6, 1995

CC CLASSIFICATION: 424

CC PRIOR APPLICATION DATA:

CC APPLICATION NUMBER: 08/350,225

CC FILING DATE: December 6, 1994

CC ATTORNEY/AGENT INFORMATION:

CC NAME: John E. Remillard

CC REGISTRATION NUMBER: 58,872

CC REFERENCE/DOCKET NUMBER: 025.6 USD2 (IM1-025CPD2)

CC TELECOMMUNICATION INFORMATION:

CC TELEPHONE: (617) 227-7430

CC TELEFAX: (617) 227-5941

CC INFORMATION FOR SEQ ID NO: 261:

CC SEQUENCE CHARACTERISTICS:

CC LENGTH: 30 amino acids

CC TYPE: amino acid

CC TOPOLOGY: linear

CC MOLECULE TYPE: peptide

CC FRAGMENT TYPE: internal

CC SEQUENCE 30 AA: 3375 MW: 4852 CN:

Query Match 46.9%; Score 97: DB 10: Length 30:  
 Best Local Similarity 56.7%; Pred. No. 2,60e-01:

Matches 14: Conservative 4: Mismatches 3: Indels 0: Gaps 0:

DB 9 SSGKTEGGNIYTKKEAFNVEK 29

QY 1 SSGKNEGNIYNNNEAFKVER 21

RESULT 11

RESULT 10

ID US-08-350-225-261 STANDARD: PRT: 30 AA.

XX xxxxxx

AC

XX

DT

XX

DE

Sequence 261, Application US/08350225

CC Sequence 261, Application US/08350225

CC GENERAL INFORMATION:

CC APPLICANT: Griffith, Irwin J.

CC APPLICANT: Pollock, Joanne;

CC APPLICANT: Bond, Julian F.

CC APPLICANT: Garman, Richard D.

CC APPLICANT: Kuo, Mei-Chang;

CC APPLICANT: Yeung, Siu-mei H.

CC APPLICANT: Brauer, Andrew;

CC APPLICANT: Exley, Mark A.

CC APPLICANT: Powers, Steven P.

CC TITLE OF INVENTION: Allergenic Proteins And Peptides From

CC TITLE OF INVENTION: Japanese Cedar Pollen

CC NUMBER OF SEQUENCES: 261

CC CORRESPONDENCE ADDRESS:

CC ADDRESSEE: Immunologic Pharmaceutical Corporation, Inc.

CC STREET: 610 Lincoln St

CC CITY: Waltham

CC STATE: MA

CC COUNTRY: USA

CC ZIP: 02154

CC COMPUTER READABLE FORM:

CC MEDIUM TYPE: Floppy disk

CC COMPUTER: IBM PC Compatible

CC OPERATING SYSTEM: PC-DOS/MS-DOS

CC SOFTWARE: Patent in Release #1.0, Version #1.25

CC CURRENT APPLICATION DATA:

CC APPLICATION NUMBER: US/08350,225

CC FILING DATE: December 6, 1994

CC CLASSIFICATION: 424

CC PRIOR APPLICATION DATA:

CC APPLICATION NUMBER: 08/226,248

CC FILING DATE: April 8, 1994

CC APPLICATION NUMBER: 07/938,990

CC FILING DATE: September 1, 1992

CC APPLICATION NUMBER: 08/0593/00139

CC FILING DATE: January 15, 1993

CC ATTORNEY/AGENT INFORMATION:

CC NAME: Carlene A. Varsone

CC REGISTRATION NUMBER: 35,729

CC REFERENCE/DOCKET NUMBER: 025.6 US (IM1-025CP2)

CC TELECOMMUNICATION INFORMATION:

CC TELEPHONE: (617) 466-6000

CC TELEFAX: (617) 466-6040

CC INFORMATION FOR SEQ ID NO: 261:

CC SEQUENCE CHARACTERISTICS:

CC LENGTH: 30 amino acids

CC TYPE: amino acid

CC TOPOLOGY: linear

CC MOLECULE TYPE: peptide

CC FRAGMENT TYPE: internal

CC SEQUENCE 30 AA: 3375 MW: 4852 CN:

Query Match 46.9%; Score 97: DB 8: Length 30:  
 Best Local Similarity 56.7%; Pred. No. 2,60e-01:

Matches 14: Conservative 4: Mismatches 3: Indels 0: Gaps 0:

DB 9 SSGKTEGGNIYTKKEAFNVEK 29

QY 1 SSGKNEGNIYNNNEAFKVER 21

```

DE US-08-467-946-261 STANDARD: PRT: 30 AA.
XX
AC xxxxxx
DE
XX
XX
DE
XX
XX
Sequence 261, Application US/08467006
Sequence 261, Application US/08467006
GENERAL INFORMATION:
APPLICANT: Griffith, Irwin J.
APPLICANT: Pollock, Joanne
APPLICANT: Bord, Julian F.
APPLICANT: Garman, Richard E.
APPLICANT: Kuo, Mei-Chang
APPLICANT: Yeung, Siu-mei H.
APPLICANT: Brauer, Andrew
APPLICANT: Exley, Mark A.
TITLE OF INVENTION: Allergic Proteins And Peptides From
Japanese Cedar Pollen
NUMBER OF SEQUENCES: 261
CORRESPONDENCE ADDRESS:
ADDRESSEE: Immunologic Pharmaceutical Corporation, Inc.
STREET: 610 Lincoln St
CITY: Waltham
STATE: MA
COUNTRY: USA
ZIP: 02154
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/467,006
FILING DATE: June 6, 1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 38/350,225
FILING DATE: December 6, 1994
ATTORNEY/AGENT INFORMATION:
NAME: Jane E. Remillard
REGISTRATION NUMBER: 36,872
REFERENCE/DOCKET NUMBER: 025.6 US06 (INT-12RCP06)
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 227-7400
TELEFAX: (617) 227-5941
INFORMATION FOR SEQ ID NO: 1
SEQUENCE 30 AA: 3375 MW: 4852 CN:
TYPE: amino acids
LENGTH: 30 amino acids
TOPOLOGY: linear
MOLECULE TYPE: Peptide
FRAGMENT TYPE: Internal
SEQUENCE CHARACTERISTICS:
SEQUENCE 30 AA: 3375 MW: 4852 CN.
Query Match 46.9% Score 97 DB 10 Length 30
Rest Local Similarity 66.7% Pred. No. 2:60e-31
Matches 14: Conservative 4: Mismatches 3: Indels 0: Gaps 0:
DB 9 SSGKVEGGNIYTKKEAFNVEK 29
QY 1 SSGKNEGNIYNNNEAFKVER 21
RESULT 12
DE US-08-467-697-261 STANDARD: PRT: 30 AA
XX
AC xxxxxx
DE
XX
XX
Sequence 261, Application US/08468940
Sequence 261, Application US/08468940
GENERAL INFORMATION:
APPLICANT: Griffith, Irwin J.
APPLICANT: Pollock, Joanne

```

```

DE
XX
XX
Sequence 261, Application US/08467697
Sequence 261, Application US/08467697
GENERAL INFORMATION:
APPLICANT: Griffith, Irwin J.
APPLICANT: Pollock, Joanne
APPLICANT: Bord, Julian F.
APPLICANT: Garman, Richard E.
APPLICANT: Kuo, Mei-Chang
APPLICANT: Yeung, Siu-mei H.
APPLICANT: Brauer, Andrew
APPLICANT: Exley, Mark A.
APPLICANT: Powers, Steven P.
TITLE OF INVENTION: Allergic Proteins And Peptides From
Japanese Cedar Pollen
NUMBER OF SEQUENCES: 261
CORRESPONDENCE ADDRESS:
ADDRESSEE: Immunologic Pharmaceutical Corporation, Inc.
STREET: 610 Lincoln St
CITY: Waltham
STATE: MA
COUNTRY: USA
ZIP: 02154
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/467,697
FILING DATE: June 6, 1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 38/350,225
FILING DATE: December 6, 1994
ATTORNEY/AGENT INFORMATION:
NAME: Jane E. Remillard
REGISTRATION NUMBER: 36,872
REFERENCE/DOCKET NUMBER: 025.6 US04 (INT-1287104)
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 227-7400
TELEFAX: (617) 227-5941
INFORMATION FOR SEQ ID NO: 1
SEQUENCE CHARACTERISTICS:
LENGTH: 30 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: Peptide
FRAGMENT TYPE: Internal
SEQUENCE 40 AA: 3375 MW: 4852 CN.
Query Match 46.9% Score 97 DB 10 Length 30
Rest Local Similarity 66.7% Pred. No. 2:60e-31
Matches 14: Conservative 4: Mismatches 3: Indels 0: Gaps 0:
DB 9 SSGKVEGGNIYTKKEAFNVEK 29
QY 1 SSGKNEGNIYNNNEAFKVER 21
RESULT 13
DE US-08-466-940-261 STANDARD: PRT: 30 AA
XX
AC xxxxxx
DE
XX
XX
Sequence 261, Application US/08468940
Sequence 261, Application US/08468940
GENERAL INFORMATION:
APPLICANT: Griffith, Irwin J.
APPLICANT: Pollock, Joanne

```

CC APPLICANT: Bond, Julian F.  
 CC APPLICANT: Garman, Richard D.  
 CC APPLICANT: Kuo, Mei-Chang  
 CC APPLICANT: Yeung, Siu-mei H.  
 CC APPLICANT: Brauer, Andrew  
 CC APPLICANT: Exley, Mark A.  
 CC APPLICANT: Powers, Steven P.  
 CC TITLE OF INVENTION: Allergenic Proteins And Peptides From  
 CC TITLE OF INVENTION: Japanese Cedar Pollen  
 CC NUMBER OF SEQUENCES: 26  
 CC CORRESPONDENCE ADDRESS:  
 CC ADDRESSEE: Immunologic Pharmaceutical Corporation, Inc.  
 CC STREET: 610 Lincoln St  
 CC CITY: Waltham  
 CC STATE: MA  
 CC COUNTRY: USA  
 CC ZIP: 02154  
 CC COMPUTER READABLE FORM:  
 CC MEDIUM TYPE: Floppy disk  
 CC COMPUTER: IBM PC Compatible  
 CC OPERATING SYSTEM: PC-DOS/MS-DOS  
 CC SOFTWARE: Patent In Release #1.0, Version #1.25  
 CC CURRENT APPLICATION DATA:  
 CC APPLICATION NUMBER: US/98/468,940  
 CC FILING DATE:  
 CC CLASSIFICATION: 424  
 CC PRIOR APPLICATION DATA:  
 CC APPLICATION NUMBER: 08/350,225  
 CC FILING DATE: December 6, 1994  
 CC APPLICATION NUMBER: 08/226,248  
 CC FILING DATE: April 8, 1994  
 CC APPLICATION NUMBER: 07/938,990  
 CC FILING DATE: September 1, 1992  
 CC APPLICATION NUMBER: PCT/US93/00139  
 CC FILING DATE: January 15, 1993  
 CC ATTORNEY/AGENT INFORMATION:  
 CC NAME: Darlene A. Vanstone  
 CC REGISTRATION NUMBER: 35,729  
 CC REFERENCE/DOCKET NUMBER: 025.6 US (IM-028CP2)  
 CC TELECOMMUNICATION INFORMATION:  
 CC TELEPHONE: (617) 466-6000  
 CC TELEFAX: (617) 466-6040  
 CC INFORMATION FOR SEQ ID NO: 261:  
 CC SEQUENCE CHARACTERISTICS:  
 CC LENGTH: 30 amino acids  
 CC TYPE: amino acid  
 CC TOPOLOGY: linear  
 CC M-RECORD TYPE: Peptide  
 CC FRAGMENT TYPE: Synthetic  
 CC SEQUENCE 30 AA: 335 MW: 4852 GN:

Query Match 45.9% Score 97 DB 10 Length 30  
 Best Local Similarity 66.7% Pred. No. 2,60e-01  
 Matches 14: Conservative 4: Mismatches 3: Indels 0: Gaps 0:

DB 9 SSGKYEGGNIYTKKEAFNVEK 29  
 IIII I III IIIIIIIII  
 QY 1 SSGKVEGTNIYNNNEAFKVER 21

RESULT 14  
 ID US-09-468-940-57 STANDARD: PRT: 20 AA  
 XX xxxxxx  
 XX  
 XX  
 XX  
 XX

Sequence 57, Application US/08468940  
 Sequence 57, Application US/08468940  
 GENERAL INFORMATION:  
 CC APPLICANT: Griffith, Irwin J.  
 CC APPLICANT: Pollock, Joanne

CC APPLICANT: Bond, Julian F.  
 CC APPLICANT: Garman, Richard D.  
 CC APPLICANT: Kuo, Mei-Chang  
 CC APPLICANT: Yeung, Siu-mei H.  
 CC APPLICANT: Brauer, Andrew  
 CC APPLICANT: Exley, Mark A.  
 CC APPLICANT: Powers, Steven P.  
 CC TITLE OF INVENTION: Allergenic Proteins And Peptides From  
 CC TITLE OF INVENTION: Japanese Cedar Pollen  
 CC NUMBER OF SEQUENCES: 26  
 CC CORRESPONDENCE ADDRESS:  
 CC ADDRESSEE: Immunologic Pharmaceutical Corporation, Inc.  
 CC STREET: 610 Lincoln St  
 CC CITY: Waltham  
 CC STATE: MA  
 CC COUNTRY: USA  
 CC ZIP: 02154  
 CC COMPUTER READABLE FORM:  
 CC MEDIUM TYPE: Floppy disk  
 CC COMPUTER: IBM PC Compatible  
 CC OPERATING SYSTEM: PC-DOS/MS-DOS  
 CC SOFTWARE: Patent In Release #1.0, Version #1.25  
 CC CURRENT APPLICATION DATA:  
 CC APPLICATION NUMBER: US/98/468,940  
 CC FILING DATE:  
 CC CLASSIFICATION: 424  
 CC PRIOR APPLICATION DATA:  
 CC APPLICATION NUMBER: 08/350,225  
 CC FILING DATE: December 6, 1994  
 CC APPLICATION NUMBER: 08/226,248  
 CC FILING DATE: April 8, 1994  
 CC APPLICATION NUMBER: 07/938,990  
 CC FILING DATE: September 1, 1992  
 CC APPLICATION NUMBER: PCT/US93/00139  
 CC FILING DATE: January 15, 1993  
 CC ATTORNEY/AGENT INFORMATION:  
 CC NAME: Darlene A. Vanstone  
 CC REGISTRATION NUMBER: 35,729  
 CC REFERENCE/DOCKET NUMBER: 025.6 US (IM-028CP2)  
 CC TELECOMMUNICATION INFORMATION:  
 CC TELEPHONE: (617) 466-6000  
 CC TELEFAX: (617) 466-6040  
 CC INFORMATION FOR SEQ ID NO: 57:  
 CC SEQUENCE CHARACTERISTICS:  
 CC LENGTH: 20 amino acids  
 CC TYPE: amino acid  
 CC TOPOLOGY: linear  
 CC M-RECORD TYPE: Peptide  
 CC FRAGMENT TYPE: Synthetic  
 CC SEQUENCE 20 AA: 221 MW: 2114 GN

Query Match 44.9% Score 93 DB 10 Length 20  
 Best Local Similarity 70.0% Pred. No. 6.84e-01  
 Matches 14: Conservative 3: Mismatches 3: Indels 0: Gaps 0:

DB 1 SSGKYEGGNIYTKKEAFNVE 20  
 IIII I IIIIIIIIIII  
 QY 1 SSGKNEGTNIYNNNEAFKVE 20

RESULT 15  
 ID US-07-938-990A-57 STANDARD: PRT: 20 AA  
 XX  
 AC xxxxxx  
 XX  
 XX  
 XX

Sequence 57, Application US/07938990A  
 Sequence 57, Application US/07938990A  
 GENERAL INFORMATION:  
 CC APPLICANT: Griffith, Irwin J.  
 CC APPLICANT: Pollock, Joanne

CC APPLICANT: Bond Julian  
 CC TITLE OF INVENTION: Allergenic Proteins And Peptides From  
 CC TITLE OF INVENTION: Japanese Cedar Pollen  
 CC NUMBER OF SEQUENCES: 70  
 CC CORRESPONDENCE ADDRESS:  
 CC ADDRESSEE: Lahive & Cockfield  
 CC STREET: Sixty State Street  
 CC CITY: Boston  
 CC STATE: MA  
 CC COUNTRY: USA  
 CC ZIP: 02109  
 CC COMPUTER READABLE FORM:  
 CC MEDIUM TYPE: Floppy disk  
 CC COMPUTER: IBM PC compatible  
 CC OPERATING SYSTEM: PC-DOS/MS-DOS  
 CC SOFTWARE: PatentIn Release #1.0, Version #1.25  
 CC CURRENT APPLICATION DATA:  
 CC APPLICATION NUMBER: US/07/935,990A  
 CC FILING DATE: 19/28/90  
 CC CLASSIFICATION: 435  
 CC PRIOR APPLICATION DATA:  
 CC APPLICATION NUMBER: 07/730,452  
 CC FILING DATE: July 15, 1991  
 CC PRIOR APPLICATION DATA:  
 CC APPLICATION NUMBER: 07/729,134  
 CC FILING DATE: July 10, 1991  
 CC ATTORNEY/AGENT INFORMATION:  
 CC NAME: Amy E. Madragouras  
 CC REGISTRATION NUMBER: 36,207  
 CC REFERENCE/DOCKET NUMBER: IPC-025CC (INT-028)  
 CC TELECOMMUNICATION INFORMATION:  
 CC TELEPHONE: (617) 227-7400  
 CC TELEFAX: (617) 227-5941  
 CC INFORMATION FOR SEQ ID NO: 57:  
 CC SEQUENCE CHARACTERISTICS:  
 CC LENGTH: 20 amino acids  
 CC TYPE: AMINO ACID  
 CC TOPOLOGY: linear  
 CC MOLECULE TYPE: peptide  
 CC FRAGMENT TYPE: internal  
 CC SEQUENCE 20 AA: 221 MW: 2174 CN:

Query Match 44.94% Score 63 DB 43 Length 20  
 Best Local Similarity 70.00% Pred. No. 6,846-91  
 Matches 14: Conservative 4 Mismatches 3 Gaps 0

DB 1 SSGKNGGNTYKKEAFNV 20  
 1111 11111111  
 0y 1 SSGKNGGNTYKKEAFNV 20

Search completed: Mon Jun 19 16:23:33 2000  
 Job time : 31 secs



\*\*\*\*\*  
 [WATERMAN] (TM)  
 \*\*\*\*\*

Release 3.1A John F. Collins, BioComputing Research Unit.  
 Copyright (C) 1993-1998 University of Edinburgh, U.K.  
 Distribution rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm  
 Run on: Mon Jun 19 16:02:46 2000; MasPar time 5.20 Seconds  
 Tabular output not generated. 281.442 Million cell updates/sec

Title: >US-09-142-524A-5  
 Description: (1-31) from US09142524A.pep  
 Perfect Score: 207  
 Sequence: 1 SSGKNEGTNIYNNNEAFKVERRF:KRVSNVI 31

Scoring table: PAM 150  
 Gap 11

Searched: 142080 seqs, 47172406 residues

Post-processing: Minimum Match 0%  
 Listing first 45 summaries

Database: pir62  
 1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 32.523; Variance 57.974; scale 0.561

Pred. No. is the number of results predicted by chance to have a  
 score greater than or equal to the score of the result being printed.  
 and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Match | Length | DP | ID     | Description             | Pred. No. |
|------------|-------|-------|--------|----|--------|-------------------------|-----------|
| 1          | 93    | 44.9  | 374    | 2  | JC2124 | major allergen Cry 1    | 2.65e-03  |
| 2          | 93    | 44.9  | 374    | 2  | JC2123 | major allergen Cry 1    | 2.65e-03  |
| 3          | 82    | 39.6  | 344    | 2  | A47025 | abiotic stress response | 1.34e-01  |
| 4          | 82    | 39.6  | 344    | 2  | S27387 | abiotic stress response | 1.34e-01  |
| 5          | 80    | 38.5  | 1101   | 2  | S66730 | hypothetical protein    | 3.14e-01  |
| 6          | 78    | 37.7  | 1308   | 2  | E71622 | probable membrane ass   | 6.32e-01  |
| 7          | 75    | 36.2  | 2380   | 2  | E71634 | hypothetical protein    | 1.77e-00  |
| 8          | 72    | 34.5  | 256    | 2  | A32017 | beta-lactamase (EC 3.   | 4.84e-00  |
| 9          | 72    | 34.8  | 260    | 1  | S11984 | fapR protein - Escher   | 4.84e-00  |
| 10         | 72    | 34.8  | 317    | 2  | B72216 | endoglucanase - Therm   | 4.84e-00  |
| 11         | 72    | 34.8  | 1355   | 2  | S51995 | probable ATPase (EC 3   | 4.84e-00  |
| 12         | 71    | 34.3  | 257    | 1  | PNBS25 | beta-lactamase (EC 3.   | 6.72e-00  |
| 13         | 71    | 34.3  | 257    | 1  | PNBS25 | beta-lactamase (EC 3.   | 6.72e-00  |
| 14         | 70    | 33.8  | 259    | 2  | I50727 | sensory organ homeobo   | 9.31e-00  |
| 15         | 70    | 33.8  | 641    | 2  | F71810 | type II: DNA modifica   | 9.31e-00  |
| 16         | 59    | 33.3  | 323    | 1  | H64130 | glycosyl transferase    | 1.29e-01  |
| 17         | 59    | 33.3  | 566    | 2  | S17477 | hypothetical protein    | 1.29e-01  |
| 18         | 59    | 33.3  | 1356   | 2  | S51389 | ROM2 protein - yeast    | 1.29e-01  |
| 19         | 58    | 32.9  | 453    | 2  | C71518 | probable permease - C   | 1.77e-01  |
| 20         | 58    | 32.9  | 610    | 2  | H71612 | asparagine--tRNA liga   | 1.77e-01  |
| 21         | 58    | 32.9  | 623    | 2  | A45050 | transketolase (EC 2.2   | 1.77e-01  |
| 22         | 58    | 32.9  | 626    | 2  | A42891 | beta-galactosidase (E   | 1.77e-01  |
| 23         | 58    | 32.9  | 929    | 2  | I51027 | type XII collagen alp   | 1.77e-01  |

24 67 32.4 158 2 F65103 PTS system, n-acetylgl 2.43e-01  
 25 67 32.4 168 2 PN0680 nitrogenase (EC 1.18. 2.43e-01  
 26 67 32.4 275 2 A69413 conserved hypothetical 2.43e-01  
 27 67 32.4 440 2 T14735 probable serine/threo 2.43e-01  
 28 67 32.4 440 2 T14735 probable serine/threo 2.43e-01  
 29 67 32.4 1234 2 B36186 I factor 2 (transposo 2.43e-01  
 30 66 31.9 169 2 S77820 probable exinuclease 1.32e-01  
 31 66 31.9 298 2 T15906 hypothetical protein 1.32e-01  
 32 66 31.9 353 2 D69001 conserved hypothetical 1.32e-01  
 33 66 31.9 393 2 A75205 bacteriochlorophyll s 1.32e-01  
 34 66 31.9 597 1 NBHUC4 C4b-binding protein a 1.32e-01  
 35 66 31.9 956 2 B71250 valine--tRNA ligase ( 1.32e-01  
 36 66 31.9 7962 2 T36345 elastic titin - human 1.32e-01  
 37 65 31.4 149 2 S14560 hemoglobin - southern 4.52e-01  
 38 65 31.4 281 2 PC5005 methionine--tRNA liga 4.52e-01  
 39 65 31.4 841 2 A41254 protein-tyrosine-phos 4.52e-01  
 40 65 31.4 1266 2 I59314 isoleucine--tRNA pho 4.52e-01  
 41 65 31.4 3433 1 GNWVKV genome polyprotein - 4.52e-01  
 42 65 31.4 3434 1 GNWVKV genome polyprotein - 4.52e-01  
 43 64 30.9 235 2 T08467 aspartate racemase (E 6.14e-01  
 44 64 30.9 1661 1 DJAD12 DNA-directed DNA poly 6.14e-01  
 45 64 30.9 1214 2 S28499 probable finger protei 6.14e-01

## ALIGNMENTS

RESULT 1 JC2124 \*type complete  
 ENTRY major allergen Cry 1 precursor (clone pCC1-15) - Japanese cedar  
 TITLE  
 ORGANISM  
 DATE  
 \*formal\_name Cryptomeria japonica \*common\_name Japanese cedar  
 14-Jul-1994 \*sequence\_revision 14-Jul-1994 \*text\_change  
 26-Aug-1999  
 JC2124  
 JC2123  
 Score: T.: Komiya, M.; Shimizu, K.; Kusakabe, T.; Morikubo, K.; Kino, K.  
 Biochem. Biophys. Res. Commun. (1994) 199:619-625  
 Cloning and sequencing of cDNA coding for Cry 1, a major allergen of Japanese cedar pollen.  
 \*cross-references MIM:94183234  
 \*accession JC2124  
 \*molecule\_type mRNA  
 \*residues 1-374 \*label: SGN  
 \*cross-references GB:D26545; NID:9493633; PID:D006087; PID:9493634  
 \*experimental\_source pollen  
 \*note the authors described carbohydrate binding site for residue 276

CLASSIFICATION #superfamily position: 1056-1066  
 KEYWORDS glycoprotein; pollen  
 FEATURE  
 1-21  
 22-374  
 \*domain signal sequence \*status predicted \*label SGN  
 \*product major allergen Cry 1 (clone pCC1-15) \*status predicted \*label MA1  
 158,191,293,354 \*binding site carbohydrate (Asn) (covalent) \*status predicted  
 SUMMARY \*length 374 \*molecular\_weight 40702 \*checksum 3692

Query Match 44.9%; Score 93; DB 2; Length 374;  
 Best Local Similarity 70.0%; Pred. No. 2.65e-03;  
 Matches 14; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Db 332 SSGKVEGGNIYTKKEAFNVE 351

QY 1 SSGKNEGTNIYNNNEAFKVE 20

RESULT 2 JC2123 \*type complete  
 ENTRY major allergen Cry 1 precursor (clone pCC1-2-2) - Japanese cedar  
 TITLE

ORGANISM  
 DATE  
 \*formal\_name Cryptomeria japonica \*common\_name Japanese cedar  
 14-Jul-1994 \*sequence\_revision 14-Jul-1994 \*text\_change



```

REFERENCE
#authors
A71600
Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.;
Aravind, L.; Koonin, E.V.; Shalloo, S.; Mason, T.; Yu, K.;
Fuji, C.; Pederson, J.; Shen, K.; Jing, J.; Aston, C.;
Lai, Z.; Schwartz, D.C.; Perle, M.; Salzberg, S.; Zhou,
L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.;
Fraser, C.M.; Adams, M.D.; Venter, J.C.; Hoffman, S.L.;
Science (1998) 282:1126-1132
Chromosome 2 sequence of the human malaria parasite
Plasmodium falciparum.
#cross-references M01D:99021743
#accession
E71600
#status
preliminary: nucleic acid sequence not shown;
translation: not shown;

##molecule_type DNA
##residues 1-1308 ##label GAR
##cross-references GB:AE001374; GB:AE001362; NID:q3845100; PID:q3845102;
TIGR:PF08125C
##experimental_source clone 307

GENETICS
#gene
PF08125C
#summary
#length 1308 #molecular-weight 155585 #checksum 5518
Query Match 37.7% Score 78; DB 2; Length 1308;
Best Local Similarity 46.2%; Pred. No. 6,32e-01;
Matches 12; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

DE 796 MNEEDFYENNETFECED:FLKREDN 821
      |||||
QY 4 KNEGTYNNNEAFKVERFIKRVSNV 29

RESULT 7
ENTRY
#type complete
#title
hypothetical protein PF0870W - malaria parasite (Plasmodium
falciparum)
#organism
#formal_name Plasmodium falciparum
#date
13-Nov-1998 #sequence-revision 13-Nov-1998 #text_change
07-May-1999
#accessions
E71604
#reference
#authors
Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.;
Aravind, L.; Koonin, E.V.; Shalloo, S.; Mason, T.; Yu, K.;
Fuji, C.; Pederson, J.; Shen, K.; Jing, J.; Aston, C.;
Lai, Z.; Schwartz, D.C.; Perle, M.; Salzberg, S.; Zhou,
L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.;
Fraser, C.M.; Adams, M.D.; Venter, J.C.; Hoffman, S.L.;
Science (1998) 282:1126-1132
Chromosome 2 sequence of the human malaria parasite
Plasmodium falciparum.
#cross-references M01D:99021743
#accession
E71604
#status
preliminary: nucleic acid sequence not shown;
translation: not shown;

##molecule_type DNA
##residues 1-2380 ##label GAR
##cross-references GB:AE001421; GB:AE001362; NID:q3845293; PID:q3845297;
TIGR:PF08070W
##experimental_source clone 307

GENETICS
#gene
PF08070W
#summary
#length 2380 #molecular-weight 283618 #checksum 2523
Query Match 36.2% Score 75; DB 2; Length 2380;
Best Local Similarity 29.2%; Pred. No. 1,77e-00;
Matches 7; Conservative 13; Mismatches 3; Indels 1; Gaps 1;

DE 122: EFLNDNLKIDKFLKKN:TIMI 1244
      |||||
QY 9 NIYNNNEAFKVERFIKRVSNV 31

RESULT 8
ENTRY
#type complete
#title
endoglucanase - Thermotoga maritima (strain MSB8)
#organism
#formal_name Thermotoga maritima

```

```

ENTRY
#type complete
#title
beta-lactamase (EC 3.5.2.6) II precursor - Bacillus cereus
(strain 57B/6)
#organism
#formal_name Bacillus cereus
#date
07-Jun-1990 #sequence-revision 07-Jun-1990 #text_change
22-Jun-1999
#accessions
A32017
#reference
A32017
Lim, H.M.; Pene, J.J.; Shaw, P.W.;
J. Bacteriol. (1988) 170:2873-2878
Cloning, nucleotide sequence, and expression of the Bacillus
cereus 57B/6 beta-lactamase II structural gene.
#cross-references M01D:88227879
#accession
A32017
##molecule_type DNA
##residues 1-256 ##label LIM
##cross-references GB:M15530; NID:q143127; PID:AA24562.1; PID:q143128
#classification
#superfamily beta-lactamase II
#keywords
antibiotic resistance; hydrolase
#feature
1-29
domain signal sequence #status predicted #label S'GN
30-256
#product beta-lactamase II #status predicted #label MA:
#length 256 #molecular-weight 28038 #checksum 682
SUMMARY
Query Match 34.8% Score 72; DB 2; Length 256;
Best Local Similarity 57.1%; Pred. No. 4,84e-00;
Matches 8; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

DB 99 MVEKKFKKRVTDVI 112
      |||||
QY 18 KVERFIKRVSNVI 31

RESULT 9
ENTRY
#type complete
#title
fapR protein - Escherichia coli
#organism
#formal_name Esch. richia coli
#date
10-Sep-1999 #sequence-revision 10-Sep-1999 #text_change
10-Sep-1999
#accessions
S11984
#reference
S11984
Klaesen, P.; de la, F.K.;
MOL. Microbiol. (1990) 4:1779-1783
Characterization of fapR, a positive regulator of expression
of the 987P operon in enterotoxigenic Escherichia coli.
#cross-references M01D:31171579
#accession
S11984
#status
preliminary
##molecule_type DNA
##residues 1-260 ##label KLA
##cross-references EMBL:X53494
#note
the sequence from Fig. 3 is inconsistent with that from
Fig. 2 in lacking 96-Lys and in having an additional
Lys after 99-Phe

GENETICS
#gene
fapR
#classification
#superfamily fapR protein
#keywords
DNA binding; transcription regulation
#summary
#length 260 #molecular-weight 30349 #checksum 6453
Query Match 34.8% Score 72; DB 1; Length 260;
Best Local Similarity 40.0%; Pred. No. 4,84e-00;
Matches 10; Conservative 6; Mismatches 9; Indels 0; Gaps 0;

DB 83 AAESYTKRSFKKRFKSCSIV 107
      |||||
QY 7 GTNYYNNNEAFKVERFIKRVSNVI 31

RESULT 10
ENTRY
#type complete
#title
endoglucanase - Thermotoga maritima (strain MSB8)
#organism
#formal_name Thermotoga maritima

```

```

DATE          11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change
11-Jun-1999
ACCESSIONS    972216
REFERENCE     A72200
#authors      Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Guinn, M.L.; Dodson,
              K.J.; Haft, D.H.; Hickey, E.K.; Peterson, J.D.; Nelson,
              W.C.; Ketchum, K.A.; McDonald, L.; Utterback, T.R.; Malek,
              J.A.; Linhorst, K.D.; Garrett, M.M.; Stewart, A.M.; Cotton,
              M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.;
              Heidelberg, J.; Sutton, G.G.; Fleischmann, R.D.; White, O.;
              Salzberg, S.L.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
#journal      Nature (1999) 399:323-329
#title        Evidence for lateral gene transfer between Archaea and
              Bacteria from genome sequence of Thermococcus maritima.
#cross-references EMBL:97287315
#accession     972216
#status        Preliminary
#molecule_type DNA
#residues      1-317 #label ARN
#cross-references GB:A5004913; GB:A500512; N1:44982321; PID:a4992329;
              TIGR:TM1751
#experimental_source strain: MS88
GENETICS      TM751
SUMMARY       #length 317 #molecular_weight 37983 #checksum 6220

Query Match      34.8%; Score 72; DB 2; Length 317;
Best Local Similarity 40.0%; Pred. No. 4,94e-00;
Matches 10; Conservative 7; Mismatches 7; Indels 1; Gaps 1;

DB 57 STRAYAPP PYK:KORFKPVDEV; 80
QY 7 GTNLYNNNEAFKVERREIKRVSNVI 3;

RESULT 1;
ENTRY   S51995 #type complete
TITLE   Probable ATPase (EC 3.6.1.-) DR52 - yeast (Saccharomyces
ALTERNATE_NAMES cerevisiae)
ORGANISM "normal_name Saccharomyces cerevisiae"
DATE 05-May-1995 #sequence_revision 01-Sep-1995 #text_change
17-Mar-1999
ACCESSIONS S51995; B54591; S30768
REFERENCE   B54596
#authors    Bussey, H.; Kaback, D.B.; Zhong, W.; Vo, D.T.; Clark, M.W.;
              Fortin, N.; Hahn, J.; Guilletto, P.F.; Keng, T.; Barton,
              A.B.; Su, V.; Davies, C.K.; Storms, R.K.
#submission Submitted to the EMBL Data Library, August 1994
#description The sequence of chromosome I of Saccharomyces cerevisiae
#accession   S51995
#molecule_type DNA
#residues    1-1355 #label BUS
#cross-references EMBL:U12980; NID:q132653; PID:a595560; MIPS:YA-0250
REFERENCE     A54591
#authors      Rippmaster, T.L.; Vaughn, G.P.; Woolford Jr., J.L.
#journal      Mol. Cell. Biol. (1993) 13:7901-7912
#title        DR51 to DR57, novel genes required for ribosome assembly and
              function in Saccharomyces cerevisiae.
#cross-references EMBL:U4067151
#accession     B54591
#molecule_type DNA
#residues      1-890; NT:893-1355 #label RIF
#cross-references EMBL:L01795
REFERENCE     S30768
#authors      Rippmaster, T.L.
#submission    Submitted to the EMBL Data Library, September 1992
#description    Identification of genes involved in ribosome assembly.
#accession     S30768
#molecule_type DNA
#residues      1-1355 #label R12
#cross-references EMBL:L01795; NID:q171113; PID:a:71114
GENETICS

```

```

#gene          SGD:DRS2; FUN18
#map_position  1L
#cross-references SGD:SGC 4; MIPS:YAL0260
FUNCTION        Probable hydrolase
KEYWORDS        hydrolase; transmembrane protein
FEATURES        251-267 #domain transmembrane #status predicted #label PKA
              448-454 #domain transmembrane #status predicted #label PKA
              498-514 #domain transmembrane #status predicted #label PKA
              1099-1115 #domain transmembrane #status predicted #label PKA
              1136-1152 #domain transmembrane #status predicted #label PKA
              1162-1178 #domain transmembrane #status predicted #label PKA
              1202-1218 #domain transmembrane #status predicted #label PKA
SUMMARY         #length 1355 #molecular_weight 154843 #checksum 7124

Query Match      34.8%; Score 72; DB 2; Length 1355;
Best Local Similarity 33.3%; Pred. No. 4,54e-00;
Matches 10; Conservative 14; Mismatches 5; Indels 0; Gaps 3;

DB 280 NNSTAEIFSEAHDDF-VEKRWII; RVGII 308
QY 4 KNEGTNIYNN-NEAFKVERREIKRVSNVI 3;

RESULT 12
ENTRY   PNBS2S #type complete
TITLE   beta-lactamase (EC 3.5.2.6) II precursor - Bacillus sp.
ALTERNATE_NAMES penicillinase II
ORGANISM "normal_name Bacillus sp."
DATE 31-Dec-1988 #sequence_revision 31-Dec-1988 #text_change
18-Jun-1999
ACCESSIONS A24393; B24393
REFERENCE   A24393
#authors    Kato, C.; Kido, T.; Watanabe, K.; Horikoshi, K.
#journal     J. Gen. Microbiol. (1985) 131:3317-3324
#title       Nucleotide sequence of the beta-lactamase gene of
              a Kalophylic Bacillus sp. strain 170.
#cross-references EMBL:86170399
#accession   A24393
#molecule_type DNA
#residues    1-257 #label KAT
#cross-references GB:M15350; NID:q:42601; PID:AAA2275 1; PID:a142602
#experimental_source strain: 170
#accession    B24393
#molecule_type protein
#residues     31-48 #label KAT2
CLASSIFICATION #superfamily bet. lactamase II
KEYWORDS       antibiotic resistance; hydrolase; metalloprotease
FEATURES        1-30 #domain signal sequence #status predicted #label SIG
              67 #product beta-lactamase II #status predicted #label BL2
              #active_site G12 #status predicted
              #binding_site zinc, high affinity (His, His, Cys, His)
              #status predicted
SUMMARY         #length 257 #molecular_weight 28153 #checksum 2375

Query Match      34.3%; Score 71; DB 1; Length 257;
Best Local Similarity 57.1%; Pred. No. 6,72e-00;
Matches 8; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

DB 100 MYEKKFKRVIDVI 113
QY 18 KVERREIKRVSNVI 31;

RESULT 13
ENTRY   PNBS02 #type complete
TITLE   beta-lactamase (EC 3.5.2.6) II precursor - Bacillus cereus
ALTERNATE_NAMES cephalosporinase II; penicillinase II
ORGANISM "normal_name Bacillus cereus"
DATE 04-Dec-1986 #sequence_revision 04-Dec-1986 #text_change
18-Jun-1999

```

```

ACCESSIONS      A91806; A91344; A01006
REFERENCE        Hussain, M.; Carlino, A.; Madonna, M.J.; Lampen, J.C.
                  J. Bacteriol. (1985) 164:223-229
                  Cloning and sequencing of the metallothioprotein
                  beta-lactamase II gene of Bacillus cereus 569/H in
                  Escherichia coli.
                  *cross-references MUID:R5C08056
                  *accession      A91806
                  *molecule_type DNA
                  *residues       1-257 *label HUS
                  *cross-references GB:M1189; NID:q142603; PIDN:AAA22276.1; PID:q142604
                  *experimental_source strain 569/H
REFERENCE        A91344
                  *authors        Amler, R.P.; Daniel, M.; Fleming, J.; Hermoso, J.M.; Pang,
                  C.; Waley, S.G.
                  *journal        FEBS Lett. (1985) 189:207-211
                  *title          The amino acid sequence of the zinc-requiring beta-lactamase
                  II from the bacterium Bacillus cereus 569.
                  *cross-references MUID:86005446
                  *accession      A91344
                  *molecule_type protein;
                  *residues       31-183;187-210;214-257 *label AMS
                  *experimental_source strain 569/H/9
COMMENT          beta-lactamase II from Bacillus spp. and beta-lactamase L-1 from
                  Pseudomonas maltophilia are classified together as class B
                  beta-lactamases, which hydrolyze cephalosporins and penicillins.
                  beta-lactamase II binds two Zn(II) ions per molecule. Zinc at the
                  higher affinity site is necessary for activity of the enzyme.
                  Zinc at the lower affinity site (not specified) increases
                  hydrolysis of cephalosporin C but not of benzylpenicillin.
CLASSIFICATION  *superfamily beta-lactamase II
KEYWORDS         antibiotic resistance; hydrolase; metalloprotein; zinc
FEATURE         1-30
                  *domain signal sequence *status predicted *label SIG\
                  *product beta-lactamase II *status experimental *label
                  MPN
                  67
                  *active_site Glu *status predicted\
                  *binding_site zinc, hlg affinity (His, His, Cys, His)
                  *status experimental
SUMMARY          *length 257 *molecular_weight 28092 *checksum 2899
Query Match     34.8% Score 711 DB 15 Length 257
Best Local Similarity 57.1% Pred. No. 6,72e+00
Matches         8: Conservative 50 Mismatches 11 Indels 0 Gaps 0
DE              109 MVEKKKQKAVDVI 113
                  :|||||:
QY              15 KVERRFKRVSNV 31
                  :|||:|||||:

RESULT 14
ENTRY   *type complete
TITLE   Sensory organ homeobox protein SOHO - chicken
ORGANISM *formal_name Gallus gallus *common_name chicken
DATE    13-Sep-1996 *sequence_revision 13-Sep-1996 *text_change
24-Sep-1999
ACCESSIONS 150727
REFERENCE 150727
          *authors    Delicher, D.L.; Pekete, D.M.; Cepko, C.L.
          *journal     J. Neurosci. (1994) 14:486-498
          *title       Asymmetric expression of a novel homeobox gene in vertebrate
          sensory organs.
          *cross-references MUID:94132877
          *accession    150727
          *status       Preliminary; translated from GB/EMBL/DBJ
          *molecule_type mRNA
          *residues     1-259 *label DEI
          *cross-references EMBL:U35815; NID:q1016799; PIDN:AAA79164.1;
          PID:q1016800
GENETICS
          *gene        S_HO-1
          *superfamily *unassigned homeobox proteins; homeobox homology

```

```

KEYWORDS        DNA binding; homeobox; nucleus; transcription regulation
FEATURE         128-184
                  *domain homeobox homology *label HOX
                  *length 259 *molecular_weight 27926 *checksum 1918
SUMMARY
Query Match     33.8% Score 701 DB 2 Length 259
Best Local Similarity 29.6% Pred. No. 9,31e+00
Matches         8: Conservative 10 Mismatches 8 Indels 1 Gaps 1
DB 124 AGCKKRTITFSQVFCLESTFDVKKR 150
QY 1 SSCKNECTIYNNNEAFKVERPF-1KR 26
                  :|||:|||||:|:|
RESULT 15
ENTRY   *type complete
TITLE   Type III DNA modification enzyme (methyltransferase);
          Helicobacter pylori (strain J99)
ORGANISM *formal_name Helicobacter pylori;
          strain J99
          *variety
DATE    12-Feb-1999 *sequence_revision 12-Feb-1999 *text_change
12-Feb-1999
ACCESSIONS F71810
REFERENCE F71810
          *authors    Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.; Brown, E.D.;
          Dolg, P.C.; Smith, D.R.; Noonan, B.; Gellid, B.C.; Delongue,
          B.L.; Carrel, G.; Tummino, P.J.; Caruso, A.;
          Uria-Nickelsen, M.; Mills, D.M.; Ives, C.; Gibson, R.;
          Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Norris,
          G.F.; Trust, T.G.
          *journal     Nature (1999) 397:176-180
          *title       Genomic sequence comparison of two unrelated isolates of the
          human gastric pathogen Helicobacter pylori.
          *cross-references MUID:99120557
          *accession    F71810
          *status       preliminary
          *molecule_type DNA
          *residues     1-641 *label ARN
          *cross-references GB:AE001563; GB:AE001439; NID:q4156027; PID:q4156030
          *experimental_source strain J99
GENETICS
          *gene        mod_2
          *length 641 *molecular_weight 72947 *checksum 9519
SUMMARY
Query Match     33.8% Score 701 DB 2 Length 641
Best Local Similarity 52.4% Pred. No. 9,31e+00
Matches         11: Conservative 13 Mismatches 5 Indels 4 Gaps 2
DB 440 SEGCGLENGUNAVFK-ERREI 459
QY 5 NEGTYNN-NEAFKVERRF 24
                  :|||:|||||:|
Search completed: Mon Jun 19 16:03:08 2000
Job time : 22 secs.

```

\*\*\*\*\*  
 W B S N E P A  
 \*\*\*\*\*  
 (TM)

Release 3.1A John F. Collins, BioComputing Research Unit.  
 Copyright (c) 1993-1998 University of Edinburgh, U.K.  
 Distribution Rights by Oxford Molecular Ltd

Mpsrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Mon Jun 19 16:01:31 2000; MasPar time 3.92 Seconds  
 Tabular output not generated. 241,115 Million cell updates/sec

File: >US-09-142-524A-5  
 Description: (1-31) from USG9142524A.pep  
 Perfect Score: 207  
 Sequence: 1 SSGKNEGTN:YNNNEAFKVERRIFKRVSNI 31

Scoring table: FAM 150  
 Gap 11

Searched: 84857 seqs, 30454973 residues

Post-processing: Minimum Match: 0%  
 Listing first 45 summaries

Database: swiss-prot38  
 1:swissprot

Statistics: Mean 33.239; Variance 51.934; scale 0.640

Pred. No. is the number of results predicted by chance to have a  
 score greater than or equal to the score of the result being printed.  
 and is derived by analysis of the total score distribution.

# SUMMARIES

| Result No. | Score | Query Match | Length | ID           | Description            | Pred. No. |
|------------|-------|-------------|--------|--------------|------------------------|-----------|
| 1          | 93    | 44.9        | 374    | 1 SBP_CRYJA  | SUGI BASIC PROTEIN PRE | 3.04e+04  |
| 2          | 82    | 39.6        | 344    | 1 ABIC_LACLA | ABORTIVE PHAGE RESIS-A | 2.89e+02  |
| 3          | 80    | 38.6        | 1101   | 1 KOES_YEAST | PROBABLE SERINE/THREON | 6.38e+02  |
| 4          | 72    | 34.8        | 256    | 1 BLAB_BACCE | BETA-LACTAMASE, TYPE I | 1.35e+00  |
| 5          | 72    | 34.8        | 1355   | 1 ATC3_YEAST | PROBABLE CALCIUM-TRANS | 1.35e+00  |
| 6          | 71    | 34.3        | 227    | 1 BLA2_BACCE | BETA-LACTAMASE, TYPE I | 1.95e+00  |
| 7          | 71    | 34.3        | 227    | 1 BLA2_BACSP | BETA-LACTAMASE, TYPE I | 1.95e+00  |
| 8          | 69    | 33.3        | 323    | 1 YF78_HAEIN | POTATIVE GLYCOSYL TRAN | 4.01e+00  |
| 9          | 69    | 33.3        | 503    | 1 CPV1_RABIT | CYTCHROME P450 19 (AR  | 4.01e+00  |
| 10         | 69    | 33.3        | 566    | 1 YCD6_YEAST | HYPOTHETICAL 64.9 KD P | 4.01e+00  |
| 11         | 69    | 33.3        | 1356   | 1 ROM2_YEAST | RHO1 GDP-GTP EXCHANGE  | 4.01e+00  |
| 12         | 68    | 32.9        | 623    | 1 TKI_HUMAN  | TRANSKETOLASE (EC 2.2. | 5.72e+00  |
| 13         | 68    | 32.9        | 626    | 1 BGAL_LEULA | BETA-GALACTOSIDASE LAR | 5.72e+00  |
| 14         | 68    | 32.9        | 929    | 1 CALC_NOTVI | COLLAGEN ALPHA 1(XII)  | 5.72e+00  |
| 15         | 67    | 32.4        | 80     | 1 Y14C_BPT4  | HYPOTHETICAL 9.5 KD PR | 8.12e+00  |
| 16         | 67    | 32.4        | 158    | 1 PTPB_ECOLI | PTS SYSTEM, N-ACETYLGA | 8.12e+00  |
| 17         | 67    | 32.4        | 211    | 1 UL92_HSV7J | PROTEIN U63            | 8.12e+00  |
| 18         | 66    | 31.9        | 113    | 1 GUAU_PIG   | UROGUANYLIN PRECURSOR  | 1.15e+01  |
| 19         | 66    | 31.9        | 353    | 1 TP6A_METH  | TYPE II DNA TOPOISOMER | 1.15e+01  |
| 20         | 66    | 31.9        | 597    | 1 C4B7_HUMAN | C4B-BINDING PROTEIN AL | 1.15e+01  |
| 21         | 66    | 31.9        | 956    | 1 SYV_TREPA  | VALYL-TRNA SYNTHETASE  | 1.15e+01  |
| 22         | 65    | 31.4        | 135    | 1 ES30_TRICO | 30 KD ANTIGENIC GLYCOP | 1.62e+01  |
| 23         | 65    | 31.4        | 149    | 1 GLB3_MCMNR | GLOBIN III.            | 1.62e+01  |

| ID | SBP_CRYJA | STANDARD | PRT  | 374 AA       |                        |          |
|----|-----------|----------|------|--------------|------------------------|----------|
| 24 | 55        | 31.4     | 281  | 1 SYM_MYCPU  | METHIONYL-TRNA SYNTHET | 1.62e+01 |
| 25 | 55        | 31.4     | 443  | 1 GLNA_PIRKO | GLUTAMINE SYNTHETASE 1 | 1.62e+01 |
| 26 | 55        | 31.4     | 497  | 1 CP26_MOUSE | CYTCHROME P450 26 (EC  | 1.14e+01 |
| 27 | 65        | 31.4     | 602  | 1 DP2_CAEEL  | PROBABLE DNA POLYMERAS | 1.62e+01 |
| 28 | 65        | 31.4     | 841  | 1 CSW_DROME  | PROTEIN-TYROSINE PHOSP | 1.62e+01 |
| 29 | 65        | 31.4     | 872  | 1 DPO1_S'LOH | DNA POLYMERASE I (EC 2 | 1.62e+01 |
| 30 | 65        | 31.4     | 1266 | 1 SYI_HUMAN  | ISOLEUCYL-TRNA SYNTHET | 1.62e+01 |
| 31 | 65        | 31.4     | 3433 | 1 POLG_KUNJM | GENOME POLYPROTEIN (CO | 1.62e+01 |
| 32 | 65        | 31.4     | 363  | 1 CAD1_STYHU | CINNAMYL-ALCOHOL DEHYD | 2.47e+01 |
| 33 | 64        | 30.9     | 389  | 1 YMF7_CAEEL | HYPOTHETICAL 45.1 KD P | 2.47e+01 |
| 34 | 64        | 30.9     | 407  | 1 AK2_BACSI  | ASPARTOKINASE 2 (EC 2. | 2.47e+01 |
| 35 | 64        | 30.9     | 432  | 1 NIFD_AZOB  | NITROGENASE MOLYBDENUM | 2.47e+01 |
| 36 | 64        | 30.9     | 500  | 1 NIFD_BRASP | NITROGENASE MOLYBDENUM | 2.47e+01 |
| 37 | 64        | 30.9     | 504  | 1 COBOL_METH | PROBABLE COBXYLIC ACID | 2.47e+01 |
| 38 | 64        | 30.9     | 623  | 1 TKI_MOUSE  | TRANSKETOLASE (EC 2.2. | 2.47e+01 |
| 39 | 64        | 30.9     | 623  | 1 TKI_RAT    | TRANSKETOLASE (EC 2.2. | 2.47e+01 |
| 40 | 64        | 30.9     | 629  | 1 CDP2_ECOLI | DITHYROLIPOAMIDE ACET  | 2.47e+01 |
| 41 | 64        | 30.9     | 891  | 1 NIA7_HORVU | NITRATE REDUCTASE (NAD | 2.47e+01 |
| 42 | 64        | 30.9     | 1061 | 1 DPO2_ADE12 | DNA POLYMERASE (EC 2.7 | 2.47e+01 |
| 43 | 64        | 30.9     | 1093 | 1 SYV_NEUCR  | VALYL-TRNA SYNTHETASE  | 2.47e+01 |
| 44 | 64        | 30.9     | 1214 | 1 TSGA_SAI   | TESTIS SPECIFIC PROTEI | 2.47e+01 |
| 45 | 63        | 30.4     | 1501 | 1 NINL_DROME | NINAC LONG PROTEIN (EC | 3.17e+01 |

## ALIGNMENTS

RESULT 1  
 ID SBP\_CRYJA STANDARD PRT: 374 AA.  
 AC P18632;  
 DT 01-NOV-1990 (Rel. 16, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE SUGI BASIC PROTEIN PRECURSOR (SBP) (MAJOR ALLERGEN CRV J 1) (CRV J 1).  
 OS Cryptomeria japonica (Japanese cedar).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC eukaryophytes; Spermatophyta; Coniferopsida; Coniferales;  
 OC Taxodiaceae; Cryptomeria  
 RN [1]  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RC TISSUE=POLLEN;  
 RX MEDLINE: 94183234.  
 RA Sone I., Komiyama N., Shimizu K., Kusakabe T., Morikubo K.,  
 RA Kino K.:  
 RT "Cloning and sequencing of cDNA coding for Crv J 1, a major allergen  
 RT of Japanese cedar pollen."  
 RL Biochem. Biophys. Res. Commun. 199 615-625(1994).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=POLLEN;  
 RA Namba M., Kurose M., Toriue K., Fukuda S., Kurimoto M.,  
 RL Submitted (JUL-1994) to the EMBL/GenBank/CDR databases.  
 RN [3]  
 RP SEQUENCE OF 22-41.  
 RC TISSUE=POLLEN;  
 RX MEDLINE: 89031257.  
 RA Tanihara M., Ando S., Usui M., Kurimoto M., Sakauchi M., Inouye S.,  
 RA Matuhara I.:  
 RT "N-terminal amino acid sequence of a major allergen of Japanese cedar  
 RT pollen (Crv J 1)."  
 RL FEBS Lett. 239:329-332(1988).  
 RN [4]  
 RP CARBOHYDRATES.  
 RX MEDLINE: 95003148.  
 RA Hijikata A., Matsumoto I., Kojima K., Ogawa H.:  
 RT "Antigenicity of the oligosaccharide moiety of the Japanese cedar  
 RT (Cryptomeria japonica) pollen allergen, Crv j1."  
 RL Int. Arch. Allergy Immunol. 105:198-202(1994).  
 CC -1- PTM: CONTAINS FUCOSE/XYLOSE-CONTAINING N-LINKED OLIGOSACCHARIDES.  
 CC -1- DISEASE: THIS PROTEIN IS THE MAJOR ALLERGEN OF JAPANESE CEDAR.  
 CC -1- POLLEN, THE MOST COMMON POLLEN ALLERGEN IN JAPAN.  
 CC -1- MISCELLANEOUS: THE SEQUENCE OF CRV J 1 FORM A IS SHOWN HERE. FORM  
 CC B DIFFERS IN SIX POSITIONS.  
 CC -1- SIMILARITY: BELONGS TO THE POLYSACCHARIDE LYASE FAMILY 1.







RC STRAIN-569/H;  
 RA MEDLINE: 8608056;  
 RA Hussain M., Carlino A., Madonna M.C., Lampen J.O.;  
 RT "Cloning and sequencing of the metalloprotein beta-lactamase II  
 gene of *Bacillus cereus* 569/H in *Escherichia coli*.";  
 J. Bacteriol. 184:223-229(1995).  
 RL [2]  
 RN [2]  
 RP SEQUENCE OF 31-183, 187-210 AND 214-257.  
 RC STRAIN-569/H;  
 RA MEDLINE: 8605446;  
 RA Ambler R.P., Balleu M., Fleming J., Herrero J.M., Pang S.G.;  
 RT "The amino acid sequence of the zinc-requiring beta-lactamase II from  
 the bacterium *Bacillus cereus* 569.";  
 FEBS Lett. 159:207-211(1985).  
 RL [3]  
 RN [3]  
 RP X-RAY CRYSTALLOGRAPHY (3.5 ANGSTROMS).  
 RC MEDLINE: 881384;  
 RA Sutton B.C., Artyukh P.J., Cordero-Sorbo A.E., Little C.,  
 RA Phillips D.C., Waley S.G.;  
 RT "An X-ray-crystallographic study of beta-lactamase II from *Bacillus*  
*cereus* at 0.45-nm resolution.";  
 Biochem. J. 242:181-188(1987).  
 RL [4]  
 RN [4]  
 RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).  
 RC MEDLINE: 9606712;  
 RA Carli A., Pares S., Duce E., Gallen M., Duce C., Frere J.-M.,  
 RA Dideberg O.;  
 RT "The 3-D structure of a zinc metallo-beta-lactamase from *Bacillus*  
*cereus* reveals a new type of protein fold.";  
 EMBO J. 14:4924-4921(1995).  
 RL [5]  
 RN [5]  
 RP X-RAY CRYSTALLOGRAPHY (1.85 ANGSTROMS).  
 RC MEDLINE: 9643752;  
 RA Carli A., Duce E., Gallen M., Frere J.-M., Dideberg O.;  
 RT "1.85-A resolution structure of the zinc (II) beta-lactamase from  
*Bacillus cereus*.";  
 Acta Crystallogr. D 54:313-323(1998).  
 RL [6]  
 RN [6]  
 RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).  
 RC MEDLINE: 99400945;  
 RA Fabiane S.M., Soti M.K., Suttan B.J.,  
 RA Padane S.M., Soti M.K., Wan T., Payne D.C., Batson J.H.,  
 RA Mitchell T., Suttan B.J.;  
 RT "Crystal structure of the zinc-dependent beta-lactamase from *Bacillus*  
*cereus* at 1.9-A resolution: binuclear active site with features of a  
 second zinc-binding site.";  
 Biochemistry 37:12404-12411(1998).  
 RL [7]  
 RN [7]  
 RP X-RAY CRYSTALLOGRAPHY (3.7 ANGSTROMS).  
 RC STRAIN-569/H;  
 RA Fabiane S.M., Soti M.K., Suttan B.J.;  
 RL Submitted (SEP-1997) to the PDB data bank.  
 CC [1]- FUNCTION: CAN HYDROLYSE CARBAPENEM COMPOUNDS.  
 CC [1]- CATALYTIC ACTIVITY: A BETA-LACTAM - H(2)O - A SUBSTITUTED BETA-  
 AMINO ACID.  
 CC [1]- COFACTOR: BINDS TWO ZINC IONS PER MOLECULE. THE ENZYME CAN ALSO  
 FUNCTION WITH ONLY ONE ZINC ION.  
 CC [1]- SUBUNIT: MONOMER.  
 CC [1]- SIMILARITY: BELONGS TO THE CLASS-B BETA-LACTAMASE FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL: M21189; AAA22276.1;  
 CC PIR: A03002; PNBSC2.  
 CC PDB: BMC; 28-AUG-96.  
 CC PDB: BMC; 07-JUL-97.  
 CC PDB: BMC; 23-SEP-98.  
 CC PDB: BMC; 14-OCT-98.

DR PDB: 2BC2; 20-APR-99.  
 DR PDB: 3BC2; 20-APR-99.  
 DR PROSITE: PS00743; BETA-LACTAMASE\_B\_1; 1.  
 DR PROSITE: PS00744; BETA-LACTAMASE\_B\_2; 1.  
 DR PFAM: PF00753; lactamase\_B;  
 KW Hydrolase; Zinc; Antibiotic resistance; Signal; 3D-structure  
 FT SIGNAL 1 30  
 FT CHAIN 31 257 BETA-LACTAMASE TYPE II  
 FT METAL 116 116 ZINC 1; HIGH-AFFINITY.  
 FT METAL 118 118 ZINC 1; HIGH-AFFINITY.  
 FT METAL 120 120 ZINC 2; LOW-AFFINITY.  
 FT METAL 179 179 ZINC 1; HIGH-AFFINITY.  
 FT METAL 198 198 ZINC 1; HIGH-AFFINITY.  
 FT METAL 240 240 ZINC 2; LOW-AFFINITY.  
 FT METAL 240 240 ZINC 2; LOW-AFFINITY.  
 SQ SEQUENCE 257 AA; 28092 MW; 268EBF87DFA45431 CR64.  
 Query Match 34.38; Score 71; DB 1; Length 257.  
 Best Local Similarity 57.13; P-Val. No. 1.95e-06;  
 Matches 8; Conservative 5; Mismatches 1; Indels 0; Gaps 0.  
 Db 100 MYKFKQKRVDCVI 113  
 Qy 18 KVERFKVSNVI 31  
 RESULT 7 STANDARD; RT: 257 AA.  
 ID BLA2\_BACSP STANDARD; RT: 257 AA.  
 AC P10425;  
 DT 01-MAR-1989 (Rel. 10, Created  
 DT 01-MAR-1989 (Rel. 10, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE BETA-LACTAMASE, TYPE II PRECURSOR (EC 3.5.2.6) (PENCILLINASE)  
 OS *Bacillus* sp. (strain 170).  
 CC Bacteria; Firmicutes; Bacilli s/Thiospirillum group;  
 CC *Bacillus/staphylococcus* group; *Bacillus*.  
 RN [1]  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 31-48.  
 RX MEDLINE: 86170399.  
 RA Kato C., Kido T., Matsumoto K., Horikoshi K.;  
 RT "Nucleotide sequence of the beta-lactamase gene of alkalophilic  
*Bacillus* sp. strain 170.";  
 J. Gen. Microbiol. 131:3317-3324(1985).  
 RL [2]  
 CC [1]- FUNCTION: CAN HYDROLYSE CARBAPENEM COMPOUNDS.  
 CC [1]- CATALYTIC ACTIVITY: A BETA-LACTAM - H(2)O - A SUBSTITUTED BETA-  
 AMINO ACID.  
 CC [1]- COFACTOR: BINDS TWO ZINC IONS PER MOLECULE.  
 CC [1]- SIMILARITY: BELONGS TO THE CLASS-B BETA-LACTAMASE FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL: M15350; AAA22275.1;  
 CC PIR: A24353; PNB52S.  
 CC HSP: P04190; LBMC.  
 DR PROSITE: PS00743; BETA-LACTAMASE\_B\_1; 1.  
 DR PROSITE: PS00744; BETA-LACTAMASE\_B\_2; 1.  
 DR PFAM: PF00753; lactamase\_B\_1.  
 KW Hydrolase; Zinc; Antibiotic resistance; Signal.  
 FT SIGNAL 1 30  
 FT CHAIN 31 257 BETA-LACTAMASE TYPE II  
 FT METAL 116 116 ZINC 1 (BY SIMILARITY).  
 FT METAL 118 118 ZINC 1 (BY SIMILARITY).  
 FT METAL 120 120 ZINC 2 (BY SIMILARITY).  
 FT METAL 179 179 ZINC 1 (BY SIMILARITY).  
 FT METAL 198 198 ZINC 2 (BY SIMILARITY).  
 FT METAL 240 240 ZINC 2 (BY SIMILARITY).  
 SQ SEQUENCE 257 AA; 28153 MW; 23EEEBDC045AE571 CR64.



Query Match: 33.3% Score 69; DB 1: Length 566;

Best Local Similarity 39.1%; Pred. No. 4; Cile-00;

Matches 9; Conservative 7; Mismatches 7; Indels 0; Gaps 0;

DB 158 DVRAKLEKVEERREKRLSDIT 180

QY 9 NLYNNKAEKVERREIKKVSNI 31

RESULT 1:

ID ROM2\_YEAST STANDARD: PRI: 1356 AA;  
AC P51862;  
DT 01-OCT-1996 (Rel. 34, Last sequence update);  
DI 15-JUL-1999 (Rel. 38, Last annotation update)  
DE RH01 GTP-EXCHANGE PROTEIN 2;  
GN ROM2 OR YLR371W OR YLR371.3;  
OS Saccharomyces cerevisiae (Baker's yeast);  
OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;  
CC Saccharomycetaceae; Saccharomycetes;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN: S288C / AB972;  
RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Du Z., Favell A., Filton D., Gattung S., Greco T., Kirsten J., Kucaba T., Hallsworth K., Hawkins J., Hillier L., Cier M., Johnson D., Johnston L., Langston Y., Latreille P., Le T., Marsden D., Mezezes S., Miller N., Nhan M., Pauley A., Peluso D., Rifkin B., Riley S., Tach A., Trevisan E., Vizzati D., Wilcox L., Wolstan P., Vaidin M., Wilson R., Wierston R.;  
RL Submitted (JAN-1995) to the EMBL/GenBank/CCSF databases;  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN: S288C / AB972;  
RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Du Z., Favell A., Filton D., Gattung S., Greco T., Kirsten J., Kucaba T., Hallsworth K., Hawkins J., Hillier L., Cier M., Johnson D., Johnston L., Langston Y., Latreille P., Le T., Marsden D., Mezezes S., Miller N., Nhan M., Pauley A., Peluso D., Rifkin B., Riley S., Tach A., Trevisan E., Vizzati D., Wilcox L., Wolstan P., Vaidin M., Wilson R., Wierston R.;  
RL Submitted (JAN-1995) to the EMBL/GenBank/CCSF databases;  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN: S288C / AB972;  
RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Du Z., Favell A., Filton D., Gattung S., Greco T., Kirsten J., Kucaba T., Hallsworth K., Hawkins J., Hillier L., Cier M., Johnson D., Johnston L., Langston Y., Latreille P., Le T., Marsden D., Mezezes S., Miller N., Nhan M., Pauley A., Peluso D., Rifkin B., Riley S., Tach A., Trevisan E., Vizzati D., Wilcox L., Wolstan P., Vaidin M., Wilson R., Wierston R.;  
RL Submitted (JAN-1995) to the EMBL/GenBank/CCSF databases;

CC Saccharomycetaceae; Saccharomycetes;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN: S288C / AB972;

RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Du Z., Favell A., Filton D., Gattung S., Greco T., Kirsten J., Kucaba T., Hallsworth K., Hawkins J., Hillier L., Cier M., Johnson D., Johnston L., Langston Y., Latreille P., Le T., Marsden D., Mezezes S., Miller N., Nhan M., Pauley A., Peluso D., Rifkin B., Riley S., Tach A., Trevisan E., Vizzati D., Wilcox L., Wolstan P., Vaidin M., Wilson R., Wierston R.;

RL Submitted (JAN-1995) to the EMBL/GenBank/CCSF databases;

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN: S288C / AB972;

RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Du Z., Favell A., Filton D., Gattung S., Greco T., Kirsten J., Kucaba T., Hallsworth K., Hawkins J., Hillier L., Cier M., Johnson D., Johnston L., Langston Y., Latreille P., Le T., Marsden D., Mezezes S., Miller N., Nhan M., Pauley A., Peluso D., Rifkin B., Riley S., Tach A., Trevisan E., Vizzati D., Wilcox L., Wolstan P., Vaidin M., Wilson R., Wierston R.;

RL Submitted (JAN-1995) to the EMBL/GenBank/CCSF databases;

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN: S288C / AB972;

RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Du Z., Favell A., Filton D., Gattung S., Greco T., Kirsten J., Kucaba T., Hallsworth K., Hawkins J., Hillier L., Cier M., Johnson D., Johnston L., Langston Y., Latreille P., Le T., Marsden D., Mezezes S., Miller N., Nhan M., Pauley A., Peluso D., Rifkin B., Riley S., Tach A., Trevisan E., Vizzati D., Wilcox L., Wolstan P., Vaidin M., Wilson R., Wierston R.;

RL Submitted (JAN-1995) to the EMBL/GenBank/CCSF databases;

RN [4]

RP SEQUENCE FROM N.A.

RC STRAIN: S288C / AB972;

RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Du Z., Favell A., Filton D., Gattung S., Greco T., Kirsten J., Kucaba T., Hallsworth K., Hawkins J., Hillier L., Cier M., Johnson D., Johnston L., Langston Y., Latreille P., Le T., Marsden D., Mezezes S., Miller N., Nhan M., Pauley A., Peluso D., Rifkin B., Riley S., Tach A., Trevisan E., Vizzati D., Wilcox L., Wolstan P., Vaidin M., Wilson R., Wierston R.;

RL Submitted (JAN-1995) to the EMBL/GenBank/CCSF databases;

RN [5]

RP SEQUENCE FROM N.A.

RC STRAIN: S288C / AB972;

RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Du Z., Favell A., Filton D., Gattung S., Greco T., Kirsten J., Kucaba T., Hallsworth K., Hawkins J., Hillier L., Cier M., Johnson D., Johnston L., Langston Y., Latreille P., Le T., Marsden D., Mezezes S., Miller N., Nhan M., Pauley A., Peluso D., Rifkin B., Riley S., Tach A., Trevisan E., Vizzati D., Wilcox L., Wolstan P., Vaidin M., Wilson R., Wierston R.;

RL Submitted (JAN-1995) to the EMBL/GenBank/CCSF databases;

RN [6]

RP SEQUENCE FROM N.A.

RC STRAIN: S288C / AB972;

RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Du Z., Favell A., Filton D., Gattung S., Greco T., Kirsten J., Kucaba T., Hallsworth K., Hawkins J., Hillier L., Cier M., Johnson D., Johnston L., Langston Y., Latreille P., Le T., Marsden D., Mezezes S., Miller N., Nhan M., Pauley A., Peluso D., Rifkin B., Riley S., Tach A., Trevisan E., Vizzati D., Wilcox L., Wolstan P., Vaidin M., Wilson R., Wierston R.;

RL Submitted (JAN-1995) to the EMBL/GenBank/CCSF databases;

RN [7]

RP SEQUENCE FROM N.A.

RC STRAIN: S288C / AB972;

RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Du Z., Favell A., Filton D., Gattung S., Greco T., Kirsten J., Kucaba T., Hallsworth K., Hawkins J., Hillier L., Cier M., Johnson D., Johnston L., Langston Y., Latreille P., Le T., Marsden D., Mezezes S., Miller N., Nhan M., Pauley A., Peluso D., Rifkin B., Riley S., Tach A., Trevisan E., Vizzati D., Wilcox L., Wolstan P., Vaidin M., Wilson R., Wierston R.;

RL Submitted (JAN-1995) to the EMBL/GenBank/CCSF databases;

RN [8]

RP SEQUENCE FROM N.A.

RC STRAIN: S288C / AB972;

RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Du Z., Favell A., Filton D., Gattung S., Greco T., Kirsten J., Kucaba T., Hallsworth K., Hawkins J., Hillier L., Cier M., Johnson D., Johnston L., Langston Y., Latreille P., Le T., Marsden D., Mezezes S., Miller N., Nhan M., Pauley A., Peluso D., Rifkin B., Riley S., Tach A., Trevisan E., Vizzati D., Wilcox L., Wolstan P., Vaidin M., Wilson R., Wierston R.;

RL Submitted (JAN-1995) to the EMBL/GenBank/CCSF databases;

RN [9]

RP SEQUENCE FROM N.A.

ID

AC

DT

DI

DE

GN

OS

OC

CC

RN

RP

RC

RA

RL

RN

RP

RC

RA

RL

RN

RP

RC

RA

RL

RN

RP

RC

RA

RL

RN

RP

RC

RA

RL

RN

RP

RC

RA

RL

RN

RP

RC

RA

RL

RN

RP

RC

RA

RL

RN

RP

RC

RA

RL

RN

RP

RC

RA

TKT\_HUMAN STANDARD: PRI: 623 AA;

P29401;

DT 01-DEC-1992 (Rel. 24, Created);

DI 01-OCT-1996 (Rel. 34, Last sequence update);

DE TRANSEKTOLOASE (EC 2.2.1.1) (TK);

GN TKT1 OR TKT;

OS Homo sapiens (Human);

OC Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;

CC Eutheria; Primates; Catarrhini; Hominoidea; Homo;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE: BRAIN;

RA McCool B.A., Plonk S.G., Martin P.K., Singleton C.K.;

RL "Cloning of human transketolase cDNAs and comparison of the

sequence of human transketolase cDNAs and comparison of the

sequence of human transketolase cDNAs and comparison of the

sequence of human transketolase cDNAs and comparison of the

sequence of human transketolase cDNAs and comparison of the

sequence of human transketolase cDNAs and comparison of the

sequence of human transketolase cDNAs and comparison of the

sequence of human transketolase cDNAs and comparison of the

sequence of human transketolase cDNAs and comparison of the

sequence of human transketolase cDNAs and comparison of the

sequence of human transketolase cDNAs and comparison of the

sequence of human transketolase cDNAs and comparison of the

sequence of human transketolase cDNAs and comparison of the

sequence of human transketolase cDNAs and comparison of the

sequence of human transketolase cDNAs and comparison of the

sequence of human transketolase cDNAs and comparison of the

sequence of human transketolase cDNAs and comparison of the

sequence of human transketolase cDNAs and comparison of the

sequence of human transketolase cDNAs and comparison of the

sequence of human transketolase cDNAs and comparison of the

sequence of human transketolase cDNAs and comparison of the

sequence of human transketolase cDNAs and comparison of the

sequence of human transketolase cDNAs and comparison of the

sequence of human transketolase cDNAs and comparison of the

sequence of human transketolase cDNAs and comparison of the

sequence of human transketolase cDNAs and comparison of the

sequence of human transketolase cDNAs and comparison of the

sequence of human transketolase cDNAs and comparison of the

sequence of human transketolase cDNAs and comparison of the

sequence of human transketolase cDNAs and comparison of the

sequence of human transketolase cDNAs and comparison of the

sequence of human transketolase cDNAs and comparison of the

sequence of human transketolase cDNAs and comparison of the

sequence of human transketolase cDNAs and comparison of the

sequence of human transketolase cDNAs and comparison of the

sequence of human transketolase cDNAs and comparison of the

sequence of human transketolase cDNAs and comparison of the

sequence of human transketolase cDNAs and comparison of the

sequence of human transketolase cDNAs and comparison of the

sequence of human transketolase cDNAs and comparison of the

sequence of human transketolase cDNAs and comparison of the

sequence of human transketolase cDNAs and comparison of the

sequence of human transketolase cDNAs and comparison of the

sequence of human transketolase cDNAs and comparison of the

sequence of human transketolase cDNAs and comparison of the

Query Match

Score 68; DB 1: Length 623;

Matches 6; Conservative 11; Mismatches 4; Indels 1; Gaps 1;

DB 706 HUPAHINDYVNNRRELKALID 727

QY 9 NLYNNKAEKVERREIKKVSNI 29

RESULT 12

Best Local Similarity 50.0%; Pred. No. 5,72e-00; Matches 9; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

DB 472 SRPENAIYNNNEQV 489  
QY 2 SKNNEGYNNEAKV 19

RESULT 13  
ID RCALIEULA STANDARD; PRT: 626 AA.  
AC Q02603;  
DT 01-JUL-1993 (Rel. 26, Created)  
DI 01-JUL-1993 (Rel. 26, Last sequence update)  
DE 15-JUL-1998 (Rel. 37, Last annotation update)  
DS BETA-GALACTOSIDASE LARGE SUBUNIT (EC 3.2.1.23) (LACTASE).  
GN LACL.  
OS Leucostoc lactis.  
OG Plasmid pN263.  
OC Bacteria: Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;  
RN Leucostoc.  
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-6.  
RX STRAIN=K25009;  
KY MEDLINE: 92325034.  
PA David S., Stevens H., van Riel M., Simons G., de Vos W.M.;  
RT "Leucostoc lactis beta-galactosidase is encoded by two overlapping genes."  
RE J. Bacteriol. 174:4475-4481(1992).  
CC 1- CATALYTIC ACTIVITY: HYDROLYSIS OF TERMINAL, NON-REDUCING BETA-D-  
CC GALACTOSE RESIDUES IN BETA-D-GALACTOSIDES.  
CC 1- SUBUNIT: HETERODIMER OF A LARGE AND A SMALL SUBUNIT.  
CC 1- SIMILARITY: BELONGS TO FAMILY 2 OF GLYCOSYL HYDROLASES.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: M92281; AAA25267.1;  
DR PIR: A42891; A42891.  
DR HSP: P00722; LBL.  
DR PRINTS: P00132; GHYDRASE2.  
DR PROSITE: PS00719; GLYCOSYL-HYDROL\_F2\_1; 1.  
DR PROSITE: PS00608; GLYCOSYL-HYDROL\_F2\_2; 1.  
DR PFAM: PF00703; Glyco\_Hydro\_2; 1.  
DR Hydrolase: Glycosidase; Plasmid.  
FT ACT\_SITE 318 318 HAS AN EFFECT ON THERMOSTABILITY  
FT ACT\_SITE 318 318 (POTENTIAL).  
FT ACT\_SITE 466 466 PROTON DONOR (BY SIMILARITY).  
FT ACT\_SITE 534 534 NUCLEOPHILE (BY SIMILARITY).  
SC SEQUENCE 626 AA: 72113 MW: E95EA1548992CF7D CRC64.  
Query Match 32.98; Score 68; DB 1; Length 626;  
Best Local Similarity 38.9%; Pred. No. 5,72e-00;  
Matches 7; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

DB 28 YHDTAEKFGSREIKSLN 45  
QY 11 YNNNEAKVREIKRVS 28

RESULT 14  
ID CALICNOTVI STANDARD; PRT: 929 AA.  
AC Q91145;  
DT 15-JUL-1998 (Rel. 36, Created)  
DI 15-JUL-1998 (Rel. 36, Last sequence update)  
DE COLLAGEN ALPHA 1(XII) CHAIN (FRAGMENT).  
DS Notophthalmus viridescens (Eastern newt) (Triturus viridescens).  
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Amphibia;  
OC

Batrachia; Caudata; Salamandridae; Salamandridae; Notophthalmus.  
RN (1)  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 95246925.  
RA Wei Y., Yang E.V., Klatt K.P., Tassava R.A.;  
RT "Monoclonal antibody MT2 identifies the urodele alpha 1 chain of type  
RT XII collagen, a developmentally regulated extracellular matrix  
RT protein in regenerating newt limbs."  
RL Dev. Biol. 168:503-513(1995).  
CC 1- FUNCTION: TYPE XII COLLAGEN INTERACTS WITH TYPE I COLLAGEN-  
CC CONTAINING FIBRILS. THE COL1 DOMAIN COULD BE ASSOCIATED WITH THE  
CC SURFACE OF THE FIBRIL. AND THE COL2 AND NC3 DOMAINS MAY BE  
CC LOCALIZED IN THE PERICELLULAR MATRIX (BY SIMILARITY). COULD PLAY  
CC A DEVELOPMENTAL ROLE IN REGENERATION.  
CC 1- SUBUNIT: TRIMER OF IDENTICAL CHAINS EACH CONTAINING 190 KD OF  
CC NONTRIPLE-HELICAL SEQUENCES (BY SIMILARITY).  
CC 1- DEVELOPMENTAL STAGE: EXPRESSION STARTS AT 3 DAYS AFTER AMPUTATION  
CC IN CELLS OF THE BASAL LAYER OF THE WOUND EPITHELIUM. AT DAY 10,  
CC EXPRESSION IS FOUND IN BOTH THE BASAL WOUND EPITHELIAL CELLS AND  
CC THE DISTAL MESENCHYME CELLS. AT MID-BOD AND LATE-BOD BLASTOMA  
CC STAGES, WOUND EPITHELIUM EXPRESSION HAS DECREASED. WHEREAS THE  
CC MESENCHYME REMAINS STRONGLY ACTIVE IN TRANSCRIPTION AND SHOWS A  
CC TENDENCY TOWARD DISTAL REGONALIZATION. CONDENSING CARTILAGE SHOWS  
CC NO SIGNAL. FINALLY, AT THE LATE DIGIT STAGE, EXPRESSION BECOMES  
CC LARGELY RESTRICTED TO THE PERICHOORDIUM.  
CC 1- PTM: THE TRIPLE-HELICAL TAIL IS STABILIZED BY DISULFIDE BONDS AT  
CC EACH END (BY SIMILARITY).  
CC 1- PTM: PROLINES AT THE THIRTY POSITION OF THE TRIPEPTIDE REPEATING  
CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS (BY  
CC SIMILARITY).  
CC 1- SIMILARITY: BELONGS TO THE FAMILY OF FIBRIL-ASSOCIATED COLLAGENS  
CC WITH INTERRUPTED HELICES (FACIT).  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: U19494; AA880217.1;  
DR HSP: P11276; IMFN.  
DR PFAM: PF00041; f-3; 7.  
DR PFAM: PF00092; vwa; 1.  
KW Extracellular matrix; Connective tissue; Repeat; Cell adhesion.  
KW Collagen; Glycoprotein.  
FT NON\_TER 1 1  
FT DOMAIN 41 52 VWFA.  
FT DOMAIN 51 154 FIBRONECTIN TYPE-III.  
FT DOMAIN 155 245 FIBRONECTIN TYPE-III.  
FT DOMAIN 246 338 FIBRONECTIN TYPE-III.  
FT DOMAIN 339 432 FIBRONECTIN TYPE-III.  
FT DOMAIN 433 519 FIBRONECTIN TYPE-III.  
FT DOMAIN 520 612 FIBRONECTIN TYPE-III.  
FT DOMAIN 613 817 VWFA.  
FT DOMAIN 818 907 FIBRONECTIN TYPE-III.  
FT DOMAIN 908 929 FIBRONECTIN TYPE-III.  
FT BINDING 231 231 TO CHONDROITIN SULFATE (POTENTIAL).  
FT BINDING 324 324 TO CHONDROITIN SULFATE (POTENTIAL).  
FT BINDING 415 415 TO CHONDROITIN SULFATE (POTENTIAL).  
FT CARBOHYD 98 98 POTENTIAL.  
FT NON\_TER 929 929  
SQ SEQUENCE 929 AA: AE5D7485254FD954 CRC64;  
Query Match 32.98; Score 68; DB 1; Length 929;  
Best Local Similarity 53.3%; Pred. No. 5,72e-00;  
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

DB 649 FKIVRNFIKRVVEVF 663  
QY 17 FKVERFIKRVNVI 31

Search completed: Mon Jun 19 16:01:52 2000  
Job time : 21 secs









DR EMBL: AFO49847; AAD30358.1; 372  
 KW Chloroplast.  
 FT NON-TER 372  
 SQ SEQUENCE 372 AA: 44233 MW: F5AE7E4F CRC32:

Query Match 34.3%; Score 71; DB 8; Length 372;  
 Best Local Similarity 35.1%; Pred. No. 4.91e+00;  
 Matches 6; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

DB 48 YONKSSLRVKKRPIFKM 24  
 QY 11 YNNNEAFKVEREIKRV 23

RESULT 11  
 ID Q9XPX5 PRELIMINARY: PR: 372 AA.

DI 01-NOV-1999 (TRENBLrel. 12, Created)  
 DT 01-NOV-1999 (TRENBLrel. 12, Last sequence update)  
 DE 01-NOV-1999 (TRENBLrel. 12, Last annotation update)  
 DE MATURASE (FRAGMENT).  
 GN MATK.  
 OS Spiridium parvifolium.  
 CC Chloroplast.  
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 CC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;  
 CC core eudicots; Rosidae; eurosids I; Rosales; Rhanaaceae; Spiridium.  
 RN [1]

RP SEQUENCE FROM N.A.  
 RC STRAIN-SPYRIPARV.  
 RA HARGIS T.M., SOLIS P.S., SOLIS D.E.  
 RT "Diversification of the North American shrub genus *Ceanothus*  
 RT (Rhamnaceae): conflicting phylogenies from nuclear DNA and  
 RT chloroplast DNA."  
 RL Am. J. Bot. 3:0-0(1999).  
 DR EMBL: AFO49849; AAD30360.1;  
 KW Chloroplast. 372 372  
 FT NON-TER  
 SQ SEQUENCE 372 AA: 44233 MW: F5AE7E4F CRC32:

Query Match 34.3%; Score 71; DB 8; Length 372;  
 Best Local Similarity 35.1%; Pred. No. 4.91e+00;  
 Matches 6; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

DB 48 YONKSSLRVKKRPIFKM 24  
 QY 11 YNNNEAFKVEREIKRV 23

RESULT 12  
 ID Q9Y2E5 PRELIMINARY: PR: 823 AA.

DI 01-NOV-1999 (TRENBLrel. 12, Created)  
 DT 01-NOV-1999 (TRENBLrel. 12, Last sequence update)  
 DE KIAA0935 PROTEIN (FRAGMENT).  
 GN KIAA0935.  
 OS Homo sapiens (Human).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 CC Eutheria; Primates; Catarrhini; Homidae; Homo.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-BRAIN.  
 RX MEDLINE: 99246063.  
 RA NAGASE T., ISH-KAWA K., SUYAMA M., KIKUNO R., HIROKAWA M.,  
 RA MIYAJIMA N., TANAKA A., KOTANI H., NOMURA N., OHARA O.  
 RT "Prediction of the coding sequences of unidentified human genes. XIII.  
 RT The complete sequences of 100 new cDNA clones from brain which code  
 RT for large proteins in vitro."  
 RL DNA Res. 6:63-76(1999).  
 DR EMBL: AB023152; BAA76779.1;  
 FT NON-TER  
 SQ SEQUENCE 823 AA: 92710 MW: BF97CB35 CRC32:

Query Match 34.3%; Score 71; DB 4; Length 823;  
 Best Local Similarity 34.8%; Pred. No. 4.91e+00;  
 Matches 9; Conservative 10; Mismatches 4; Indels 1; Gaps 1;

DB 598 IYSDANGYQMORRPVSYVNSI 520  
 QY 10 IYNNNEAFKVEREIKRVSNV 31

RESULT 13  
 ID Q91964 PRELIMINARY: PR: 259 AA.

DI 01-NOV-1999 (TRENBLrel. 01, Created)  
 DT 01-NOV-1999 (TRENBLrel. 01, Last sequence update)  
 DE 01-NOV-1999 (TRENBLrel. 12, Last annotation update)  
 DE SENSORY ORGAN HOMEOBOX PROTEIN S.O.H.O.  
 GN SOHO-1.  
 OS Gallus gallus (Chicken).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Archosauria; Anser.  
 CC Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 94132877.  
 RA DEITCHER D.L., PEKETE D.M., CEPKO G.L.  
 RT "Asymmetric expression of a novel homeobox gene in vertebrate sensory  
 RT organs."  
 RL J. Neurosci. 14:486-498(1994).  
 RN [2]

RP SEQUENCE FROM N.A.  
 RA DEITCHER D.L.  
 RT Submitted (SEP-1995) to the EMBL/Genbank/DBJ databases.  
 CC -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).  
 DR EMBL: J35815; AAT79164.1;  
 DR EMBL: S69380; AAC60747.1;  
 DR HSSP: P22808; LNK3.  
 DR PROSITE: PS00027; HOMEOD 1; 1.  
 DR PFAM: PF00046; homeobox .  
 DR PRINTS: PR00024; HOMEOBOX.  
 KW Homeobox; DNA-binding; Nuclear protein.  
 SQ SEQUENCE 259 AA: 27926 MW: 03819495 CRC32:

Query Match 33.8%; Score 70; DB 13; Length 259;  
 Best Local Similarity 29.6%; Pred. No. 6.99e+00;  
 Matches 11; Conservative 11; Mismatches 4; Indels 1; Gaps 1;

DB 124 AGGKKKRTIFSKSVFQENSHVKKP 190  
 QY 1 SSGKNEGIMYNNNEAFKVEREIKRV 20

RESULT 14  
 ID Q03561 PRELIMINARY: PR: 505 AA.

DI 01-JUL-1997 (TRENBLrel. 04, Created)  
 DT 01-AUG-1998 (TRENBLrel. 07, Last sequence update)  
 DE 01-NOV-1999 (TRENBLrel. 12, Last annotation update)  
 DE MATURASE.  
 GN YCF14 OR MATK.  
 OS Cardenia thunbergia.  
 CC Chloroplast.  
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 CC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;  
 CC core eudicots; Asteridae; euasterids I; Gentianales; Rubiaceae;  
 CC Cardenia.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA ENDRESS M.E., SENNBAD B., NILSSON S., CIVEYREL L., CHASE M.W.,  
 RA HUYSMANS S., GRAFSTROEM E., BREMER B.  
 RL Opera Bot. Belg. 7:59-102(1996).  
 DR EMBL: Z7C198; CAA94087.1;  
 DR MENDE : 23592; Garth:Ycf14;23592  
 DR PFAM: PF01348; Intron\_matur\_s2; 1.



\*\*\*\*\*  
 WIREH  
 \*\*\*\*\* (14)

Release 4.1A John F. Collins, BioComputing Research Unit.  
 Copyright (c) 1993-1998 University of Edinburgh, U.K.  
 Distribution Rights by Oxford Molecular Ltd

XPsearch\_pp Protein - protein database search, using Smith-Waterman algorithm

Run on: Mon Jun 19 16:19:11 2000: MasPar time 2.63 Seconds  
 Tabular output not generated. 31.065 Million cell updates/sec

Title: >US-09-142-524A-14  
 Description: (1-9), from US09142524A.pep  
 Perfect Score: 57  
 Sequence: 1 SIXRVSNVI 9

Scoring table: PAM 150  
 Gap 15

Searched: 14963 seqs, 24636106 residues

Post-processing: Minimum Match 04  
 Listing first 45 summaries

Database: a-geneseq45  
 l-geneseqf

Statistics: Mean 15.426, Variance 44.241, scale 0.349

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score | Query Match | Length | ID     | Description                 | Pred. No. |
|------------|-------|-------------|--------|--------|-----------------------------|-----------|
| 1          | 53    | 43.0        | 9      | W57751 | Residues 109-117 of Cry J 1 | 1.87e-01  |
| 2          | 53    | 43.0        | 15     | W57751 | Residues 109-120 of Cry J 1 | 1.87e-01  |
| 3          | 53    | 43.0        | 15     | R89231 | Japanese cedar pollen       | 1.87e-01  |
| 4          | 53    | 43.0        | 20     | R45552 | Cry J 1 pollen allergen     | 1.87e-01  |
| 5          | 53    | 43.0        | 20     | R45552 | Cry J 1 Japanese cedar      | 1.87e-01  |
| 6          | 53    | 43.0        | 30     | W44563 | T-cell epitope peptide      | 1.87e-01  |
| 7          | 53    | 43.0        | 31     | W27372 | Multi-epitope peptide       | 1.87e-01  |
| 8          | 53    | 43.0        | 31     | W27372 | Multi-epitope peptide       | 1.87e-01  |
| 9          | 53    | 43.0        | 31     | W27372 | Multi-epitope peptide       | 1.87e-01  |
| 10         | 53    | 43.0        | 305    | W27372 | Multi-epitope peptide       | 1.87e-01  |
| 11         | 53    | 43.0        | 134    | W27372 | Multi-epitope peptide       | 1.87e-01  |
| 12         | 53    | 43.0        | 353    | R75339 | Japanese cedar pollen       | 1.87e-01  |
| 13         | 53    | 43.0        | 353    | R81597 | Cedar pollen allergen       | 1.87e-01  |
| 14         | 53    | 43.0        | 374    | R60166 | Japanese cedar pollen       | 1.87e-01  |
| 15         | 53    | 43.0        | 374    | R31937 | Cry J 1                     | 1.87e-01  |
| 16         | 53    | 43.0        | 374    | R82440 | Cry J 1 Japanese cedar      | 1.87e-01  |
| 17         | 53    | 43.0        | 374    | R25541 | Cry J 1 pollen allergen     | 1.87e-01  |
| 18         | 46    | 80.7        | 20     | R25541 | Cry J 1 Japanese cedar      | 1.20e-02  |
| 19         | 46    | 80.7        | 20     | R45552 | Cry J 1 pollen allergen     | 1.20e-02  |
| 20         | 44    | 77.2        | 500    | R72593 | Acromonium crysogenum       | 2.02e-02  |
| 21         | 44    | 77.2        | 566    | W86348 | Human multiple endocrin     | 2.02e-02  |
| 22         | 44    | 77.2        | 610    | W29719 | Tumor suppressor prote      | 2.02e-02  |
| 23         | 42    | 73.7        | 292    | R04895 | Phellinase Insulin          | 3.35e-02  |

|    |    |      |      |   |        |                         |          |
|----|----|------|------|---|--------|-------------------------|----------|
| 24 | 42 | 73.7 | 641  | 2 | W11473 | Marek's disease virus   | 4.35e-02 |
| 25 | 41 | 71.9 | 171  | 1 | W20206 | H. pylori derived prot  | 4.31e-02 |
| 26 | 41 | 71.9 | 266  | 1 | W98839 | H. pylori GHPD 1533 pr  | 4.31e-02 |
| 27 | 41 | 71.9 | 276  | 1 | W20776 | H. pylori flagella-ass  | 4.31e-02 |
| 28 | 41 | 71.9 | 343  | 1 | W90011 | Expressed antigen for   | 4.31e-02 |
| 29 | 41 | 71.9 | 357  | 1 | R46665 | Rat RCL7 serotonergic   | 4.31e-02 |
| 30 | 41 | 71.9 | 357  | 1 | R46665 | Murine SH75A serotonin  | 4.31e-02 |
| 31 | 41 | 71.9 | 367  | 1 | W51279 | Babesia microti HMA1-3  | 4.31e-02 |
| 32 | 41 | 71.9 | 367  | 1 | W56301 | Babesia microti antigen | 4.31e-02 |
| 33 | 41 | 71.9 | 401  | 1 | W93395 | S. erythraea dDp-D-6-d  | 4.31e-02 |
| 34 | 41 | 71.9 | 401  | 1 | W19741 | Sugar biosynthesis enz  | 4.31e-02 |
| 35 | 41 | 71.9 | 445  | 1 | W15558 | Helicobacter polypteri  | 4.31e-02 |
| 36 | 41 | 71.9 | 445  | 1 | W88793 | H. pylori GHPD 1252 pr  | 4.31e-02 |
| 37 | 41 | 71.9 | 452  | 1 | W20598 | HPV1: mutant L1 protei  | 4.31e-02 |
| 38 | 41 | 71.9 | 501  | 1 | W20599 | Human brx protein.      | 4.31e-02 |
| 39 | 41 | 71.9 | 1429 | 1 | W93941 | Melanocortin-4 recepto  | 5.54e-02 |
| 40 | 40 | 70.2 | 332  | 1 | W97624 | Melanocortin-4 recepto  | 5.54e-02 |
| 41 | 40 | 70.2 | 332  | 1 | W95687 | Human MC4 protein.      | 5.54e-02 |
| 42 | 40 | 70.2 | 332  | 1 | W92442 | PP2 protein.            | 5.54e-02 |
| 43 | 40 | 70.2 | 360  | 1 | W98912 | PP2 disease resistant   | 5.54e-02 |
| 44 | 40 | 70.2 | 917  | 1 | R80132 | CER protein.            | 5.54e-02 |
| 45 | 40 | 70.2 | 1054 | 1 | W30923 |                         |          |

#### ALIGNMENTS

RESULT 1  
 ID W57751 standard; peptide: 9 AA.  
 AC W57751:  
 DE 17-SEP-1998 (first entry)  
 DE Residues 109-117 of Cry J 1  
 KW Cry J 1, Japanese cedar pollen antigen; allergy: immunotherapy;  
 KW HLA class II molecule.  
 CS Cryptomeria japonica.  
 PN WC9820402-A1.  
 PD 22-MAY-1998.  
 PF 12-NOV-1997: J04329.  
 PR 13-NOV-1996: JP-302053.  
 PA (MEP) MEIJI MILK PROD CO L  
 PI Daifuku K. Kuno, K. Kume A, T.  
 DR WP: 99-297817/26.  
 PR Peptides derived from Japan cedar pollen antigens are  
 PT immunotherapeutic agents useful for allergy treatment and typing  
 PT HLA class II molecules in allergy sufferers  
 PS Claim 12: Page 26; 50pp; Japanese.  
 CC This sequence represents residues 109-117 of the Cry J 1 protein, and  
 CC is a peptide of the invention. The peptides are derived from Japanese  
 CC cedar pollen antigens, and are used as immunotherapeutic agents in the  
 CC treatment of allergy. The peptides can be used for identification, and  
 CC typing of the particular HLA class II molecules in an allergy sufferer,  
 CC and also for peptide immunotherapy of an allergy. Using these peptides,  
 CC the immunotherapy can be targeted more specifically to the recipients  
 CC of the individual patient, allowing more effective treatment of an  
 CC allergy, including those patients for whom treatment with a conventional  
 CC immunotherapeutic agent is ineffective.  
 SQ Sequence 9 AA:

Query Match 93.0% Score 53: DB 1: length 9:  
 Best Local Similarity 100.0% Pred. No. 1.87e-01:  
 Matches 8: Conservative C: Mismatches 0: Indels 0: Gaps 0:

Db 2 IKRVSNVI 9  
 QY 2 IKRVSNVI 9

RESULT 2  
 ID W57750 standard; peptide: 15 AA.  
 AC W57750:  
 DE 17-SEP-1998 (first entry)  
 DE Residues 106-120 of Cry J 1  
 KW Cry J 1, Japanese cedar pollen antigen; allergy: immunotherapy;  
 KW HLA class II molecule.

US Cryptomeria japonica.  
 PD W09820402-A1.  
 PE 22-MAY-1998.  
 PF 12-NOV-1997: JP4129.  
 PR 12-NOV-1996: JP-302053.  
 PA (MEIP) ; MEIP: MILK PROD CO LTD.  
 PI Dairiki, K, Kino K, Kume A, Sone T.  
 UR WPI: 98-297517/26.  
 PT Peptides derived from Japanese cedar pollen antigens are  
 PT immunotherapeutic agents - useful for allergy treatment and typing  
 PI HLA class II molecules in allergy sufferers  
 PS Claim 12: Page 26: Supp: Japanese.  
 CC This sequence represents residues 106-120 of the Cry j 1 protein, and  
 CC is a peptide of the invention. The peptides are derived from Japanese  
 CC cedar pollen antigens, and are used as immunotherapeutic agents in the  
 CC treatment of allergy. The peptides can be used for identification and  
 CC typing of the particular HLA class II molecules in an allergy sufferer,  
 CC and also for peptide immunotherapy of an allergy. Using these peptides,  
 CC the immunotherapy can be targeted more specifically to the requirements  
 CC of the individual patient, allowing more effective treatment of an  
 CC allergy, including those patients for whom treatment with a conventional  
 CC immunotherapeutic agent is ineffective.  
 CC Sequence 15 AA:

Query Match 93.0% Score 53: DB 1: Length 15:  
 Best Local Similarity 100.0% Pred. No. 1.87e-01:  
 Matches 8: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

Db 5 IKRVSNVI 12  
 IIILIIII  
 QY 2 IKRVSNVI 4  
 RESULT 3  
 ID R89291 standard; peptide: 15 AA.  
 AC R89291.  
 DE 12-MAY-1996 (first entry)  
 PE Japanese cedar pollen allergen (Cry) I derived 1-cell epitope peptide.  
 KW Japanese cedar pollen allergen (Cry) I 1-cell epitope: peptides:  
 KW prevention; treatment; cryptomeria pollen: sis.  
 PS Cryptomeria japonica.  
 PI J6716285-A  
 PD 05-MAY-1995  
 PF 20-OCT-1993: 262626.  
 PR 20-OCT-1993: JP-262626.  
 PA (MEIP) ; MEIP: MILK PROD CO LTD.  
 UR WPI: 95-203834/27.  
 PT New cryptomeria pollen allergen 1-cell epitope peptide - used for  
 PT prevention; treatment and investigation of Japanese cedar pollens  
 PS Claim 5: Page 2: Supp: Japanese.  
 CC R75388 is the Japanese cedar pollen allergen Cry I, from which the  
 CC 1-cell epitope peptides R89291-R89295 were derived. The peptides  
 CC can be used for the prevention and treatment of cryptomeria  
 CC pollens, and also for the investigation of pollens.  
 CC Sequence 15 AA:

Query Match 93.0% Score 53: DB 1: Length 15:  
 Best Local Similarity 100.0% Pred. No. 1.87e-01:  
 Matches 8: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

Db 5 IKRVSNVI: 12  
 IIILIIII  
 QY 2 IKRVSNVI 4  
 RESULT 4  
 ID R45552 standard; Protein: 20 AA.  
 AC R45552.  
 DE 13-JUL-1994 (first entry)  
 PE Cry j 1 pollen allergen peptide CJI-11.  
 KW Japanese cedar: detection; allergy; treatment; diagnosis:  
 KW 1-cell epitope; sensitivity.  
 PS Cryptomeria japonica.

PN W09401560-A.  
 PD 20-JAN-1994.  
 PE 15-JAN-1993: U00139.  
 PR 10-JUL-1992: WO-056661.  
 PA (IMMU) ; IMMUNOLOGIC PHARM CORP.  
 PI Bond JF, Garman RD, Griffith LL, Kuo M, Pollock J.  
 UR WPI: 94-035666/94.  
 PT Antigens derived from Japanese cedar pollen allergen Cry j 1  
 PT contain at least two T cell epitope(s), used to treat or diagnose  
 PT allergy  
 PS Claim 1: Fig 13: 137pp: English.  
 CC The sequence is that of an isolated peptide of the Japanese cedar  
 CC pollen allergen Cry j 1 (amino acids 101-120). The peptide CJI-11  
 CC can be used for the treatment and diagnosis of allergies associated  
 CC with Japanese cedar pollen. It has enhanced therapeutic properties  
 CC but reduced side effects compared to naturally occurring allergens  
 CC Sequence 20 AA:

Query Match 93.0% Score 53: DB 1: Length 20:  
 Best Local Similarity 100.0% Pred. No. 1.87e-01:  
 Matches 8: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

Db 10 IKRVSNVI 17  
 IIILIIII  
 QY 2 IKRVSNVI 9  
 RESULT 5  
 ID R42501 standard; Protein: 20 AA.  
 AC R42501.  
 DE 15-APR-1996 (first entry)  
 PE Cry j 1 Japanese Cedar pollen allergen peptide fragment (CJI-11).  
 KW Cry j 1; Japanese cedar pollen allergen; modified; drug production;  
 KW allergy; Cryptomeria japonica.  
 OS Cryptomeria japonica.  
 PN W09527786-A1.  
 PD 19-OCT-1995.  
 PF 06-APR-1995: U04249  
 PR 06-APR-1994: US-226248.  
 PA (IMMU) ; IMMUNOLOGIC PHARM CORP.  
 PI Chen X, Evans S, Franzen BM, Kuo M, Powers SP.  
 UR Shaked Z.  
 WPI: 95-366391/47.  
 PT Modified cryptomeria japonica (Cry j 1) peptides - useful for  
 PT treating allergy to Japanese cedar pollen allergen or  
 PT immunologically cross reacting allergens  
 PS Disclosures: Figure 2: Supp: English.  
 CC Novel peptides of Cry j 1 have been modified as a part of a  
 CC proteinization scheme to develop an optimised drug product for  
 CC therapeutic treatment of humans suffering from allergy to Japanese  
 CC cedar pollen allergen or an allergen which is immunologically cross  
 CC reactive with Japanese cedar pollen allergen. Such modified peptides  
 CC possess certain characteristics which render them particularly  
 CC suitable for drug product formulation. Peptide fragments of Cry j 1  
 CC modified and unmodified, are given in R82491-R82525. This peptide  
 CC fragment corresponds to amino acids 101-120 of the allergen mature  
 CC protein.  
 CC Sequence 20 AA:

Query Match 93.0% Score 53: DB 1: Length 20:  
 Best Local Similarity 100.0% Pred. No. 1.87e-01:  
 Matches 8: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

Db 10 IKRVSNVI 17  
 IIILIIII  
 QY 2 IKRVSNVI 9  
 RESULT 6  
 ID W44683 standard; peptide: 30 AA.  
 AC W44683.

```

DI 01-MAY-1996 (first entry)
DE T-cell epitope peptide #2 of sugi pollen antigen.
KW T-cell epitope; sugi pollen antigen; sugi pollinosis.
OS Synthetic.
OS Crytomoria japonica.
PN 21007700-A.
PD 13-JAN-1998.
PF 24-JUN-1996; 163287.
PA (DAI) : DAICEL CHEM IND LTD.
PA (MEI) : MEIJI SEIKA KAKIASHA LTD.
DR WPI: 97-470495/43.
PT T-cell epitope peptide of sugi pollen antigen - useful in the
  treatment of sugi pollinosis.
PS Claim 1; Page 4; 14pp; Japanese.
CC T-cell epitope peptides W4582-88 and their derivatives react with
  sugi pollinosis patient peripheral blood T lymphocytes. A composition
  prepared by combining at least 2 of the above peptides and/or their
  derivatives is used for the prevention and treatment of sugi pollinosis.
SQ Sequence 31 AA.

Query Match 93.0% Score 53; DB 1; Length 31;
Best Local Similarity 100.0%; Pred. No. 1.87e-01;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 10 IKRVSNVI 17
      |||||
QY 2 IKRVSNVI 9

RESULT 7
ID W27372 standard; peptide: 31 AA.
AC W27372.
DI 24-MAR-1998 (first entry)
DE Multi-epitope peptide used as immunotherapeutic agent #4.
KW Multi-epitope peptide; immunotherapeutic agent; allergic disease;
  T-cell epitope region; allergen; lymphocyte; immunoglobulin E.
OS Synthetic.
OS Crytomoria japonica.
PN W2732600-A1.
PD 12-SEP-1997.
PF 10-MAR-1996; 100740.
PA (MEI) : MEIJI MILK PROD CO LTD.
PA Dairiki K, Iwama A, Kino K, Kume A, Sone T.
DR WPI: 97-470495/43.
PT Peptide immunotherapeutic agent to treat allergic diseases.
  contains multi-epitope peptide containing T cell epitope regions
  from different allergens.
PS Claim 5; Page 32; 58pp; Japanese.
CC The present sequence represents a multi-epitope peptide which is used as
  a new immunotherapeutic agent. It comprises T cell epitope regions from 2
  or more different allergens (preferably linked via arginine or lysine
  dimers), where the T cell epitope regions have a positivity index
  greater than 100 as measured in a patient group responding to the
  allergen. Have at least 70% reactivity with lymphocytes from patients
  (IgE) antibodies from patients responsive to the allergen. The agent can
  be used to prevent and treat a wide variety of allergic diseases, e.g. by
  desensitisation. Side effects, e.g. those mediated by IgE, are reduced.
SQ Sequence 31 AA.

Query Match 93.0% Score 53; DB 1; Length 31;
Best Local Similarity 100.0%; Pred. No. 1.87e-01;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 24 IKRVSNVI 31
      |||||
QY 2 IKRVSNVI 9

RESULT 8
ID W27373 standard; peptide: 31 AA.
AC W27373.
DI 24-MAR-1998 (first entry)
DE Multi-epitope peptide used as immunotherapeutic agent #5.
KW Multi-epitope peptide; immunotherapeutic agent; allergic disease;
  T-cell epitope region; allergen; lymphocyte; immunoglobulin E.
OS Synthetic.
OS Crytomoria japonica.
PN W2732600-A1.
PD 12-SEP-1997.
PF 10-MAR-1996; 100740.
PA (MEI) : MEIJI MILK PROD CO LTD.
PA Dairiki K, Iwama A, Kino K, Kume A, Sone T.
DR WPI: 97-470495/43.
PT Peptide immunotherapeutic agent to treat allergic diseases.
  contains multi-epitope peptide containing T cell epitope regions
  from different allergens.
PS Claim 5; Page 32; 58pp; Japanese.
CC The present sequence represents a multi-epitope peptide which is used as
  a new immunotherapeutic agent. It comprises T cell epitope regions from 2
  or more different allergens (preferably linked via arginine or lysine
  dimers), where the T cell epitope regions have a positivity index
  greater than 100 as measured in a patient group responding to the
  allergen. Have at least 70% reactivity with lymphocytes from patients
  (IgE) antibodies from patients responsive to the allergen. The agent can
  be used to prevent and treat a wide variety of allergic diseases, e.g. by
  desensitisation. Side effects, e.g. those mediated by IgE, are reduced.
SQ Sequence 31 AA.

Query Match 93.0% Score 53; DB 1; Length 31;
Best Local Similarity 100.0%; Pred. No. 1.87e-01;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 24 IKRVSNVI 31
      |||||
QY 2 IKRVSNVI 9

RESULT 9
ID W27369 standard; peptide: 80 AA.
AC W27369.
DI 24-MAR-1998 (first entry)
DE Multi-epitope peptide used as immunotherapeutic agent #1.
KW Multi-epitope peptide; immunotherapeutic agent; allergic disease;
  T-cell epitope region; allergen; lymphocyte; immunoglobulin E.
OS Synthetic.
OS Crytomoria japonica.
PN W2732600-A1.
PD 12-SEP-1997.
PF 10-MAR-1996; 100740.
PA (MEI) : MEIJI MILK PROD CO LTD.
PA Dairiki K, Iwama A, Kino K, Kume A, Sone T.
DR WPI: 97-470495/43.
PT Peptide immunotherapeutic agent to treat allergic diseases.
  contains multi-epitope peptide containing T cell epitope regions
  from different allergens.
PS Claim 6; Page 31; 58pp; Japanese.
CC The present sequence represents a multi-epitope peptide which is used as
  a new immunotherapeutic agent. It comprises T cell epitope regions from 2
  or more different allergens (preferably linked via arginine or lysine
  dimers), where the T cell epitope regions have a positivity index
  greater than 100 as measured in a patient group responding to the
  allergen. Have at least 70% reactivity with lymphocytes from patients
  (IgE) antibodies from patients responsive to the allergen. The agent can
  be used to prevent and treat a wide variety of allergic diseases, e.g. by
  desensitisation. Side effects, e.g. those mediated by IgE, are reduced.
SQ Sequence 80 AA.

Query Match 93.0% Score 53; DB 1; Length 80;
Best Local Similarity 100.0%; Pred. No. 1.87e-01;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 18 IKRVSNVI 25

```

```

DI 24-MAR-1998 (first entry)
DE Multi-epitope peptide used as immunotherapeutic agent #5.
KW Multi-epitope peptide; immunotherapeutic agent; allergic disease;
  T-cell epitope region; allergen; lymphocyte; immunoglobulin E.
OS Synthetic.
OS Crytomoria japonica.
PN W2732600-A1.
PD 12-SEP-1997.
PF 10-MAR-1996; 100740.
PA (MEI) : MEIJI MILK PROD CO LTD.
PA Dairiki K, Iwama A, Kino K, Kume A, Sone T.
DR WPI: 97-470495/43.
PT Peptide immunotherapeutic agent to treat allergic diseases.
  contains multi-epitope peptide containing T cell epitope regions
  from different allergens.
PS Claim 9; Page 32; 58pp; Japanese.
CC The present sequence represents a multi-epitope peptide which is used as
  a new immunotherapeutic agent. It comprises T cell epitope regions from 2
  or more different allergens (preferably linked via arginine or lysine
  dimers), where the T cell epitope regions have a positivity index
  greater than 100 as measured in a patient group responding to the
  allergen. Have at least 70% reactivity with lymphocytes from patients
  (IgE) antibodies from patients responsive to the allergen. The agent can
  be used to prevent and treat a wide variety of allergic diseases, e.g. by
  desensitisation. Side effects, e.g. those mediated by IgE, are reduced.
SQ Sequence 31 AA.

Query Match 93.0% Score 53; DB 1; Length 31;
Best Local Similarity 100.0%; Pred. No. 1.87e-01;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 24 IKRVSNVI 31
      |||||
QY 2 IKRVSNVI 9

```

```

Query Match          93.0%  Score 53: DB 1: Length 134:
Best Local Similarity 100.0%: Pred. No. 1: 87e-01:
Matches      8: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

DB 18 IKRVSNVI 25
      |||||
QY 2 IKRVSNVI 9

RESULT 12
ID R75388 standard: protein; 353 AA.
AC R75388:
DE 12-MAR-1996 (first entry)
KW Japanese cedar pollen allergen Cryj I.
KW Japanese cedar pollen allergen; Cryj I; T-cell epitope; peptides-
KW prevention; treatment; cryptomeria pollinosis.
OS Cryptomeria japonica.
FH Key Location/Qualifiers
FT peptide 61..75
FT peptide /note= "T-cell epitope peptide"
FT peptide 91..105
FT peptide /note= "T-cell epitope peptide"
FT peptide 105..120
FT peptide /note= "T-cell epitope peptide"
FT peptide 146..160
FT peptide /note= "T-cell epitope peptide"
FT peptide 211..225
FT peptide /note= "T-cell epitope peptide"
FT peptide 326..340
FT peptide /note= "T-cell epitope peptide"
FT peptide 335..346
FT peptide /note= "T-cell epitope peptide"
PN J07118295-A.
PD 09-MAY-1995.
PF 20-OCT-1993: 262626.
PR 20-OCT-1993: JP-262626.
PA (MEIP ) MEIJI MILK PROD CO LTD.
ER WPI: 95-203834/27.
PT New cryptomeria pollen allergen T-cell epitope peptide - used for
PT prevention, treatment and investigation of Japanese cedar pollinosis.
PS Disclosure; Figs 1-2: epp; Japanese.
DS R75388 is the Japanese cedar pollen allergen Cryj I, from which the
CC T-cell epitope peptides R89289-R89295 were derived. The peptides
CC can be used for the prevention and treatment of cryptomeria
CC pollinosis, and also for the investigation of pollinosis.
SQ Sequence 353 AA:

Query Match          93.0%  Score 53: DB 1: Length 134:
Best Local Similarity 100.0%: Pred. No. 1: 87e-01:
Matches      8: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

DB 110 IKRVSNVI 117
      |||||
QY 2 IKRVSNVI 9

RESULT 13
ID R81587 standard: Protein; 353 AA.
AC R81587:
DE 24-MAY-1996 (first entry)
KW Cedar pollen allergen B.
KW Cedar pollen allergen; immunoglobulin E; IgE; T-cell epitope;
KW antibody; pollinosis; therapy; immunotherapy.
OS Cryptomeria japonica.
PN EP-700929-A2.
PD 13-MAR-1996.
PF 08-SEP-1995: 306295.
PR 10-SEP-1994: JP-242137.
PR 14-JUL-1995: JP-200221.
PR 14-JUL-1995: JP-200204.
PA (HAYB ) HAYASHIBARA SEIBUTSU KAGAKU.

```

PI Hino K., Saito S., Taniguchi Y.  
 DR WPI: 96-140276/15.  
 PT New peptides derived from cedar pollen allergens - activate  
 PI allergen-specific T-cells, but not allergen-specific IgE antibodies.  
 PI used for treating cedar pollinosis.  
 PS Claim 5: Page 31-32; 36pp; English.  
 CC Synthetic peptides based on portions of cedar pollen allergens A  
 CC (R81586) and B (R81587) were tested for their ability to activate  
 CC cedar allergen-specific T-cells, but not allergen-specific IgE  
 CC antibodies. 5 peptides (R81580-R81585) were identified as T-cell  
 CC epitopes. These peptides, plus subsequences (R81573-79) essential  
 CC for T-cell recognition, and homologous peptides (R81588-96) can  
 CC be used as immunotherapeutic agents to treat or prevent cedar  
 CC pollinosis, avoiding side-effects such as anaphylaxis.  
 SQ Sequence 374 AA.

Query Match 93.0% Score 53 DB 17 Length 374  
 Best Local Similarity 100.0% Pred No. 1.87e+01

Matches 8: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

DB 131 IKRVSNI: 132

QY 2 IKRVSNI: 9

RESULT 14  
 IC R60164 standard: Protein: 374 AA.  
 AC R60164

DT 24-MAR-1995 (first entry)

DE Japanese cedar pollen antigen CryjI.

KW Japanese cedar: pollen antigen; allergen; CryjI: sig1: pollinosis.

CS Cryptomeria japonica.

FR Key Location/Qualifiers

FI peptide 1..21

FI protein 22..374

FI label= signal\_peptide

FI label= mature\_CryjI

PN JCM97768-A.

PD 19-JUL-1994

PF 07-JAN-1993: G01116.

PR 07-JAN-1993: JP-C01116.

PA (MEL) MEL: SEIKA KAISHA.

DR WPI: 94-263680/23.

DR N-PSDB: Q71501.

PI Sugi (Japanese cedar) pollen antigen CryjI - is useful for

PI diagnosis, treatment and prevention of sugi pollinosis

PS Claim 2: Page 5-7; 9pp; Japanese.

CC The coding sequence for the Japanese cedar ("sugi") pollen allergen

CC CryjI was isolated from a cDNA library prepared from polyA mRNA. All

CC or part of the CryjI protein can be used for diagnosis treatment

CC and prevention of sugi pollinosis.

SQ Sequence 374 AA.

Query Match 93.0% Score 53 DB 17 Length 374  
 Best Local Similarity 100.0% Pred No. 1.87e+01

Matches 8: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

DB 131 IKRVSNI: 138

QY 2 IKRVSNI: 9

RESULT 15

IC R31937 standard: Protein: 374 AA.

AC R31937

DT 03-JUN-1993 (first entry)

DE Cry j I.

KW Japanese cedar pollen: allergen; antigen; allergy; B cell; T cell.

CS Cryptomeria japonica.

FR Key Location/Qualifiers

FI peptide 1..21

FI label= "signal peptide"

FI protein 22..374

PT Hino K., Saito S., Taniguchi Y.  
 DR WPI: 96-140276/15.  
 PT New peptides derived from cedar pollen allergens - activate  
 PI allergen-specific T-cells, but not allergen-specific IgE antibodies.  
 PI used for treating cedar pollinosis.  
 PS Claim 5: Page 31-32; 36pp; English.  
 CC Synthetic peptides based on portions of cedar pollen allergens A  
 CC (R81586) and B (R81587) were tested for their ability to activate  
 CC cedar allergen-specific T-cells, but not allergen-specific IgE  
 CC antibodies. 5 peptides (R81580-R81585) were identified as T-cell  
 CC epitopes. These peptides, plus subsequences (R81573-79) essential  
 CC for T-cell recognition, and homologous peptides (R81588-96) can  
 CC be used as immunotherapeutic agents to treat or prevent cedar  
 CC pollinosis, avoiding side-effects such as anaphylaxis.  
 SQ Sequence 374 AA.

Query Match 93.0% Score 53 DB 17 Length 374  
 Best Local Similarity 100.0% Pred No. 1.87e+01

Matches 8: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

DB 131 IKRVSNI: 138

QY 2 IKRVSNI: 9

PN JCM97768-A.

PD 19-JUL-1994

PF 07-JAN-1993: G01116.

PR 07-JAN-1993: JP-C01116.

PA (MEL) MEL: SEIKA KAISHA.

DR WPI: 94-263680/23.

DR N-PSDB: Q71501.

PI Sugi (Japanese cedar) pollen antigen CryjI - is useful for

PI diagnosis, treatment and prevention of sugi pollinosis

PS Claim 2: Page 5-7; 9pp; Japanese.

CC The coding sequence for the Japanese cedar ("sugi") pollen allergen

CC CryjI was isolated from a cDNA library prepared from polyA mRNA. All

CC or part of the CryjI protein can be used for diagnosis treatment

CC and prevention of sugi pollinosis.

SQ Sequence 374 AA.

Query Match 93.0% Score 53 DB 17 Length 374  
 Best Local Similarity 100.0% Pred No. 1.87e+01

Matches 8: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

DB 131 IKRVSNI: 138

QY 2 IKRVSNI: 9

Search completed: Mon Jun 19 16:19:22 2000

Job time : 11 secs.

\*\*\*\*\*

(TM)

\*\*\*\*\*

Release 3.1A John F. Collins, BioComputing Research Unit,  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

Mpsrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Tue Jun 20 13:37:46 2000; Maspar time 2.17 Seconds

Tabular output not generated. 59,945 Million cell updates/sec

Title: >US-09-142-524A-14  
Description: (1-9) from US05142524A.pep  
Perfect Score: 57  
Sequence: 1 SKRYSNV: 9

Scoring table: FAM 150  
Gap 15

Searched: 145341 seqs, 14437480 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: a-issued

1:5A\_COMB 2:5B\_COMB 3:6\_COMB 4:PCT\_COMB 5:backfiles1

Statistics: Mean 14.393; Variance 44.43; scale 0.324

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Query Match % | Length | LP | ID         | Description            | Pred. No. |
|------------|-------|---------------|--------|----|------------|------------------------|-----------|
| 1          | 44    | 77.2          | 564    | 2  | US-08-565- | Sequence 1, Applicatio | 1.39e+02  |
| 2          | 44    | 77.2          | 610    | 2  | US-08-565- | Sequence 6, Applicatio | 1.39e+02  |
| 3          | 42    | 73.7          | 255    | 1  | US-08-459- | Sequence 4, Applicatio | 2.27e+02  |
| 4          | 42    | 73.7          | 255    | 1  | US-08-459- | Sequence 4, Applicatio | 2.27e+02  |
| 5          | 41    | 71.9          | 357    | 1  | US-08-356- | Sequence 2, Applicatio | 2.90e+02  |
| 6          | 41    | 71.9          | 357    | 1  | US-08-356- | Sequence 4, Applicatio | 2.90e+02  |
| 7          | 41    | 71.9          | 401    | 2  | US-08-576- | Sequence 45, Applicati | 2.90e+02  |
| 8          | 41    | 71.9          | 445    | 2  | US-08-679- | Sequence 32, Applicati | 2.90e+02  |
| 9          | 41    | 71.9          | 2544   | 2  | US-08-576- | Sequence 32, Applicati | 2.90e+02  |
| 10         | 40    | 70.2          | 332    | 2  | US-08-780- | Sequence 8, Applicatio | 3.69e+02  |
| 11         | 40    | 70.2          | 332    | 2  | US-08-780- | Sequence 8, Applicatio | 3.69e+02  |
| 12         | 40    | 70.2          | 332    | 2  | US-08-842- | Sequence 8, Applicatio | 3.69e+02  |
| 13         | 40    | 70.2          | 332    | 2  | US-08-842- | Sequence 8, Applicatio | 3.69e+02  |
| 14         | 40    | 70.2          | 332    | 1  | US-08-671- | Sequence 8, Applicatio | 3.69e+02  |
| 15         | 40    | 70.2          | 885    | 4  | PCT-US95-0 | Sequence 2, Applicatio | 3.69e+02  |
| 16         | 40    | 70.2          | 885    | 4  | PCT-US95-0 | Sequence 2, Applicatio | 3.69e+02  |
| 17         | 40    | 70.2          | 885    | 2  | PCT-US95-0 | Sequence 2, Applicatio | 3.69e+02  |
| 18         | 40    | 70.2          | 909    | 4  | PCT-US95-0 | Sequence 142, Applicat | 3.69e+02  |
| 19         | 40    | 70.2          | 909    | 2  | US-08-310- | Sequence 142, Applicat | 3.69e+02  |
| 20         | 39    | 68.4          | 33     | 4  | PCT-US95-0 | Sequence 97, Applicati | 4.68e+02  |
| 21         | 39    | 68.4          | 33     | 4  | PCT-US95-0 | Sequence 97, Applicati | 4.68e+02  |
| 22         | 39    | 68.4          | 33     | 2  | US-08-310- | Sequence 97, Applicati | 4.68e+02  |
| 23         | 39    | 68.4          | 227    | 1  | US-08-745- | Sequence 6, Applicatio | 4.68e+02  |

|    |    |      |      |   |            |                         |          |
|----|----|------|------|---|------------|-------------------------|----------|
| 24 | 39 | 68.4 | 227  | 1 | US-08-597- | Sequence 6, Applicatio  | 4.68e+02 |
| 25 | 39 | 68.4 | 317  | 2 | US-08-518- | Sequence 2, Applicatio  | 4.68e+02 |
| 26 | 39 | 68.4 | 317  | 2 | US-08-518- | Sequence 2, Applicatio  | 4.68e+02 |
| 27 | 39 | 68.4 | 317  | 3 | US-08-518- | Sequence 2, Applicatio  | 4.68e+02 |
| 28 | 39 | 68.4 | 346  | 1 | US-08-518- | Sequence 4, Applicatio  | 4.68e+02 |
| 29 | 39 | 68.4 | 503  | 4 | PCT-US95-0 | Sequence 8, Applicatio  | 4.68e+02 |
| 30 | 39 | 68.4 | 503  | 1 | US-08-307- | Sequence 8, Applicatio  | 4.68e+02 |
| 31 | 39 | 68.4 | 503  | 1 | US-08-474- | Sequence 8, Applicatio  | 4.68e+02 |
| 32 | 39 | 68.4 | 503  | 1 | US-08-245- | Sequence 8, Applicatio  | 4.68e+02 |
| 33 | 39 | 68.4 | 1019 | 2 | US-08-222- | Sequence 26, Applicatio | 4.68e+02 |
| 34 | 39 | 68.4 | 1019 | 1 | US-08-271- | Sequence 7, Applicatio  | 4.68e+02 |
| 35 | 38 | 66.7 | 12   | 1 | US-07-778- | Sequence 34, Applicati  | 5.94e+02 |
| 36 | 38 | 66.7 | 12   | 1 | US-07-778- | Sequence 34, Applicati  | 5.94e+02 |
| 37 | 38 | 66.7 | 192  | 1 | US-08-085- | Sequence 80, Applicati  | 5.94e+02 |
| 38 | 38 | 66.7 | 192  | 4 | US-08-085- | Sequence 80, Applicati  | 5.94e+02 |
| 39 | 38 | 66.7 | 585  | 4 | US-08-085- | Sequence 80, Applicati  | 5.94e+02 |
| 40 | 38 | 66.7 | 641  | 1 | US-08-151- | Sequence 1, Applicatio  | 5.94e+02 |
| 41 | 38 | 66.7 | 641  | 1 | US-08-151- | Sequence 1, Applicatio  | 5.94e+02 |
| 42 | 38 | 66.7 | 681  | 2 | US-08-555- | Sequence 4, Applicatio  | 5.94e+02 |
| 43 | 38 | 66.7 | 681  | 4 | PCT-US96-0 | Sequence 4, Applicatio  | 5.94e+02 |
| 44 | 38 | 66.7 | 803  | 5 | US-08-745- | Sequence 4, Applicatio  | 5.94e+02 |
| 45 | 38 | 66.7 | 807  | 4 | PCT-US96-0 | Sequence 2, Applicatio  | 5.94e+02 |

## ALIGNMENTS

|        |   |          |      |         |
|--------|---|----------|------|---------|
| RESULT | 1   | STANDARD | PAT: | 564 AA. |
| ID     | US-08-565-337A-1  |          |      |         |
| XX     | xxxxxx  |          |      |         |
| AC     |   |          |      |         |
| XX     |   |          |      |         |
| DI     |   |          |      |         |
| XX     |   |          |      |         |
| DE     | Sequence 1, Application US/-8865337A                        |          |      |         |
| XX     |   |          |      |         |
| CC     | Sequence 1, Application US/-8865337A                        |          |      |         |
| CC     | Patent No. 5972649  |          |      |         |
| CC     | GENERAL INFORMATION:  |          |      |         |
| CC     | APPLICANT: Ac-Young, Janice                                 |          |      |         |
| CC     | APPLICANT: Covitz, Peter                                    |          |      |         |
| CC     | APPLICANT: Tang, Y. Tom                                     |          |      |         |
| CC     | APPLICANT: Murry, Lyng                                      |          |      |         |
| CC     | TITLE OF INVENTION: HUMAN MULTIPLE ENDONUCLEASE PLASMA TYPE |          |      |         |
| CC     | TITLE OF INVENTION: PROTEIN                                 |          |      |         |
| CC     | NUMBER OF SEQUENCES: 6                                      |          |      |         |
| CC     | CORRESPONDENCE ADDRESS:                                     |          |      |         |
| CC     | ADDRESSEE: Inocyte Pharmaceuticals Inc.                     |          |      |         |
| CC     | STREET: 3174 Porter Drive                                   |          |      |         |
| CC     | CITY: Palo Alto   |          |      |         |
| CC     | STATE: CA   |          |      |         |
| CC     | COUNTRY: USA  |          |      |         |
| CC     | ZIP: 94304  |          |      |         |
| CC     | COMPUTER READABLE FORM:                                     |          |      |         |
| CC     | MEDIUM TYPE: Diskette                                       |          |      |         |
| CC     | COMPUTER: IBM Compatible                                    |          |      |         |
| CC     | OPERATING SYSTEM: DOS                                       |          |      |         |
| CC     | SOFTWARE: FastSeq for Windows Version 2.0                   |          |      |         |
| CC     | CURRENT APPLICATION DATA:                                   |          |      |         |
| CC     | APPLICATION NUMBER: US/08/865,337A                          |          |      |         |
| CC     | FILING DATE: Herewith                                       |          |      |         |
| CC     | CLASSIFICATION: 530   |          |      |         |
| CC     | PRIOR APPLICATION DATA:                                     |          |      |         |
| CC     | APPLICATION NUMBER:   |          |      |         |
| CC     | FILING DATE:  |          |      |         |
| CC     | ATTORNEY/AGENT INFORMATION:                                 |          |      |         |
| CC     | NAME: Billings, Lucy  |          |      |         |
| CC     | REGISTRATION NUMBER: 749                                    |          |      |         |
| CC     | REFERENCE/DOCKET NUMBER: PF-0305 US                         |          |      |         |
| CC     | TELECOMMUNICATION INFORMATION:                              |          |      |         |
| CC     | TELEPHONE: 415-855-0555                                     |          |      |         |
| CC     | TELEFAX: 415-845-4166                                       |          |      |         |
| CC     | TELEX:  |          |      |         |
| CC     | INFORMATION FOR SEQ ID NO: 1:                               |          |      |         |



SEQUENCE CHARACTERISTICS:  
LENGTH: 564 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE: LIBRARY BRAINOTII

CLONE: Consensus  
SEQUENCE 564 AA: 62110 MW: 1662460 CN:

Query Match 77.2% Score 44: DB 2: Length 564:  
Best Local Similarity 52.5% Pred. No. 1339c-02:  
Matches 5: Conservative 3: Mismatches 0: Indels 0: Gaps 0:

DB 118 VKKSVLVI 125  
QY 2 IKRVSNV 9

RESULT 2  
ID US-06-HES-337A-4 STANDARD: PRI: 610 AA.  
XX xxxxxx  
AC  
CT  
DT  
ET  
FT  
GT  
HT  
IT  
JT  
KT  
LT  
MT  
NT  
OT  
PT  
QT  
RT  
ST  
TT  
UT  
VT  
WT  
XT  
YT  
ZT

Sequence 5: Application US/08865337A  
Sequence 6: Application US/08865337A  
Patent No. 5972549  
GENERAL INFORMATION:  
APPLICANT: Au-Young, Janice  
APPLICANT: Cavilz, Peter  
APPLICANT: Tang, Y. Tom  
APPLICANT: Xue, J. Lynn E.  
TITLE OF INVENTION: HUMAN MULTIPLE ENDOCRINE NEOPLASIA TYPE  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Inveit Pharmaceuticals, Inc.  
Street: 3174 Porter Drive  
City: Palo Alto  
State: CA  
Country: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: Fastseq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
FILING DATE: Heposide  
CLASSIFICATION: 51C  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 35,749  
REFERENCE/DOCKET NUMBER: PF-0305 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-845-0555  
TELEFAX: 415-845-4156  
TELEX:  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 510 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE: LIBRARY GenBank

CC CLONE: 1945387  
SQ SEQUENCE 610 AA: 67466 MW: 1878916 CN:  
Query Match 77.2% Score 44: DB 2: Length 610:  
Best Local Similarity 62.5% Pred. No. 1339c-02:  
Matches 5: Conservative 3: Mismatches 0: Indels 0: Gaps 0:

DB 118 VKKSEVL 125  
QY 2 IKRVSNV 9

RESULT 3  
ID US-08-459-264-4 STANDARD: PRI: 255 AA.  
XX xxxxxx  
AC  
CT  
DT  
ET  
FT  
GT  
HT  
IT  
JT  
KT  
LT  
MT  
NT  
OT  
PT  
QT  
RT  
ST  
TT  
UT  
VT  
WT  
XT  
YT  
ZT

Sequence 4: Application US/0445-264  
Sequence 4: Application US/ 459264  
Patent No. 5705340  
GENERAL INFORMATION:  
APPLICANT: RASMUSSEN, Bette A  
APPLICANT: TALLY, Francis P  
APPLICANT: GLUZMAN, Yakov  
TITLE OF INVENTION: DNA LIGASES AND AMINO ACIDS SEQUENCES  
TITLE OF INVENTION: CLASS B BETA-LACTAMASE ENZYME FROM PATTERIES  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Parby & Darryl  
Street: 805 Third Avenue  
City: New York  
State: NY  
Country: US  
ZIP: 10022  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Research #1.0, Version #1.0  
CURRENT APPLICATION DATA:  
FILING DATE: US/08/459,264  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Patricia A. Sells  
REGISTRATION NUMBER: 44,719  
REFERENCE/DOCKET NUMBER: 44,719  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-527-7700  
TELEFAX: 212-753-6237  
TELEX: 236687  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 255 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
ORIGINAL SOURCE:  
ORGANISM: Bacillus cereus  
IMMEDIATE SOURCE:  
CLONE: B-LACTAMASE  
SEQUENCE 255 AA: 27847 MW: 349604 CN:

Query Match 73.7% Score 42: DB 2: Length 255:  
Best Local Similarity 71.4% Pred. No. 2,27e-02:  
Matches 5: Conservative 2: Mismatches 0: Indels 0: Gaps 0:

DB 105 KKVTVI 111

```

QY      IIIII:
      3 KRVSNI 9
RESULT  4
ID      US-08-459-263-4      STANDARD:      PRT:      255 AA.
XX      XXXXXX
XX
XX
DE      Sequence 4, Application US/08459263
XX      Sequence 4, Application US/08459263
XX      Patent No. 5792642
XX      GENERAL INFORMATION:
XX      APPLICANT: RASMUSSEN, Beth A
XX      APPLICANT: TALLY, Francis P
XX      APPLICANT: GLUZMAN, Yakov
XX      TITLE OF INVENTION: DNA SEQUENCES AND AMINO ACID SEQUENCES
XX      TITLE OF INVENTION: OF CLASS B BETA-LACTAMASE ENZYMES FROM BACTERIOIDES
XX      TITLE OF INVENTION: FRAGILIS
XX      NUMBER OF SEQUENCES: 9
XX      CORRESPONDENCE ADDRESS:
XX      ADDRESSEE: Darby & Darby PC
XX      STREET: 805 Third Avenue
XX      CITY: New York
XX      STATE: NY
XX      COUNTRY: US
XX      ZIP: 10022
XX      COMPUTER READABLE FORM:
XX      MEDIUM TYPE: Floppy disk
XX      COMPUTER: IBM PC compatible
XX      OPERATING SYSTEM: PC-DOS/MS-DOS
XX      SOFTWARE: Patent In Release #1.0, Version #1.30
XX      CURRENT APPLICATION DATA:
XX      APPLICATION NUMBER: US/08/459,263
XX      FILING DATE: 02-JUN-1995
XX      CLASSIFICATION: 435
XX      ATTORNEY/AGENT INFORMATION:
XX      NAME: Robinson, Joseph R
XX      REGISTRATION NUMBER: 33,448
XX      REFERENCE/DOCKET NUMBER: 0646/1B026-US2
XX      TELECOMMUNICATION INFORMATION:
XX      TELEPHONE: 212-527-7700
XX      TELEFAX: 212-753-6237
XX      TELEX: 234687
XX      INFORMATION FOR SEQ ID NO: 4:
XX      SEQUENCE CHARACTERISTICS:
XX      LENGTH: 255 amino acids
XX      TYPE: amino acid
XX      STRANDEDNESS: single
XX      TOPOLOGY: linear
XX      MOLECULE TYPE: peptide
XX      ORIGINAL SOURCE:
XX      ORGANISM: Bacillus cereus
XX      IMMEDIATE SOURCE:
XX      CLONE: B-LACTAMASE
XX      SEQUENCE 255 AA: 27647 MW: 349604 CN:

Query Match 73.7% Score 42: DB 1: Length 255:
Best Local Similarity 71.4% Pred. NO. 2,278-02:
Matches 5: Conservative 2: Mismatches 0: Indels 0: Gaps 0:

DB 105 KRVIDV: IIII:
QY      IIIII:
      3 KRVSNI 9
RESULT  5
ID      US-08-356-405-2      STANDARD:      PRT:      357 AA.
XX      XXXXXX
XX
XX
DE      Sequence 4, Application US/08031538
XX      Sequence 4, Application US/08031538
XX      Patent No. 5968817
XX
XX
DE      Sequence 2, Application US/08356405
XX      Sequence 2, Application US/08356405
XX      Patent No. 5807691
XX      GENERAL INFORMATION:
XX      APPLICANT: Anlaiky, No. 5807691:rdine
XX      APPLICANT: Boschert, Ursula
XX      APPLICANT: Hen, Rene
XX      APPLICANT: Plassat, Jean-Luc
XX      TITLE OF INVENTION: Polypeptides Having Serotonin Receptor
XX      TITLE OF INVENTION: Activity (SHT5A), Nucleic Acids Coding for These
XX      TITLE OF INVENTION: Polypeptides and Uses Thereof:
XX      NUMBER OF SEQUENCES: 13
XX      CORRESPONDENCE ADDRESS:
XX      ADDRESSEE: Rhone-Poulenc Rorer Inc.
XX      STREET: 500 Arcola Road, 3043
XX      CITY: Collegeville
XX      STATE: PA
XX      COUNTRY: USA
XX      ZIP: 19426-0107
XX      COMPUTER READABLE FORM:
XX      MEDIUM TYPE: Floppy disk
XX      COMPUTER: IBM PC compatible
XX      OPERATING SYSTEM: PC-DOS/MS-DOS
XX      SOFTWARE: Patent In Release #1.0, Version #1.25
XX      CURRENT APPLICATION DATA:
XX      APPLICATION NUMBER: US/08/356,405
XX      FILING DATE:
XX      CLASSIFICATION: 435
XX      PRIOR APPLICATION DATA:
XX      APPLICATION NUMBER: WO PCT/EP91/00650
XX      FILING DATE: 29-JUN-1992
XX      PRIOR APPLICATION DATA:
XX      APPLICATION NUMBER: FR 92/06091
XX      FILING DATE: 01-JUL-92
XX      ATTORNEY/AGENT INFORMATION:
XX      NAME: Smith, Julie K
XX      REGISTRATION NUMBER: 19,615
XX      REFERENCE/DOCKET NUMBER: EX92004-US
XX      TELECOMMUNICATION INFORMATION:
XX      TELEPHONE: (610)454-3849
XX      TELEFAX: (610)454-3808
XX      INFORMATION FOR SEQ ID NO: 2:
XX      SEQUENCE CHARACTERISTICS:
XX      LENGTH: 357 amino acids
XX      TYPE: amino acid
XX      TOPOLOGY: linear
XX      MOLECULE TYPE: protein
XX      SEQUENCE 357 AA: 43804 MW: 741957 CN:

Query Match 71.9% Score 41: DB 1: Length 357:
Best Local Similarity 85.7% Pred. NO. 2,506-02:
Matches 6: Conservative 1: Mismatches 0: Indels 0: Gaps 0:

DB 155 KRVSVM 151
QY      IIIII:
      3 KRVSNI 9
RESULT  6
ID      US-08-031-538-4      STANDARD:      PRT:      357 AA.
XX      XXXXXX
XX
XX
DE      Sequence 4, Application US/08031538
XX      Sequence 4, Application US/08031538
XX      Patent No. 5968817
XX
XX

```



[illegible]

ery Match 20 28. score

TELEX: 32:  
INFORMATION FOR SEO ID NC: 32:



```

CC FILING DATE:
CC CLASSIFICATION: 516
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Smith, Deann F.
CC REGISTRATION NUMBER: 36683
CC REFERENCE/DOCKET NUMBER: 2115-000853DVB
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (810)641-1600
CC TELEFAX: (810)641-0270
CC INFORMATION FOR SEQ ID NO: 8:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 332 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC SEQ SEQUENCE 332 AA; 36956 MW; 598933 CN;

Query Match 70.2% Score 40; DB 2; Length 332;
Rest Local Similarity 44.4%; Pred. No. 3,696-02;
Matches 4; Conservative 4; Mismatches 1; Indels 0;

DB: 162 IVKRGV:II 170
QY : IIII:
      1 SIRVSNI: 9

RESULT 14
ID US-671-525B-8 STANDARD: PRI: 332 AA.
XX xxxxxx
XX
XX
CX
XX
DE Sequence 8, Application US/08671525B
XX Sequence 8, Application US/08671525B
CX Patent No. 5703220
CC GENERAL INFORMATION:
CC APPLICANT: Yamada, Tadataka
CC TITLE OF INVENTION: Genes Encoding Melanocortin Receptors
CC NUMBER OF SEQUENCES: 23
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Harness, Dickey & Pierce, P.L.C.
CC STREET: P.O. Box 828
CC CITY: Bloomfield Hills
CC STATE: MI
CC COUNTRY: US
CC ZIP: 48303
CC COMPUTER READABLE FORM
CC MEDIUM TYPE: floppy disk
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patent Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/671,525B
CC FILING DATE: June 27, 1996
CC CLASSIFICATION: 435
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Smith, Deann F.
CC REGISTRATION NUMBER: 36683
CC REFERENCE/DOCKET NUMBER: 2115-000853DVB
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (810)641-1600
CC TELEFAX: (810)641-0270
CC INFORMATION FOR SEQ ID NO: 8:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 332 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC SEQ SEQUENCE 332 AA; 36956 MW; 598933 CN;

```

```

Query Match          70.2%   Score 40: DB 1: Length 332:
Best Local Similarity 44.4%   Pred.No. 3.69e-02:
Matches           4: Conservative      Mismatches 1: Indels

Db    162 TVKRVGII:170
QY     1 SIKRVSNI 9

RESULT 15
ID PCT-US95-34S89-2 STANDARD: FRT: 885 AA.
XX
AC xxxxxx
AD
DT
DX
DE Sequence 2, Application: PC/US9504589
CC
CC Sequence 2, Application PC/US9504589
CC GENERAL INFORMATION:
CC APPLICANT: Ausubel, Frederick M.
CC APPLICANT: Staskawicz, Brian J.
CC APPLICANT: Brent, Andrew F.
CC APPLICANT: Dabiback, Douglas
CC APPLICANT: Katagiri, Fumiki
CC APPLICANT: Kunkel, Barbara N.
CC APPLICANT: Mindrins, Michael N.
CC APPLICANT: Yu, Guo-Liang
CC TITLE OF INVENTION: RPS GENE AND USES THEREOF
CC NUMBER OF SEQUENCES: 20
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Fish & Richardson
CC STREET: 225 Franklin Street Suite 3100
CC CITY: Boston
CC STATE: MA
CC COUNTRY: USA
CC ZIP: 02110-2904
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC OPERATING SYSTEM: IBM PC Compatible
CC SOFTWARE: Patent Release #1.0, Version #1.30B
CC CURRENT APPLICATION DATA: PC/US95/34589
CC APPLICATION NUMBER: PC/US95/34589
CC FILING DATE:
CC CLASSIFICATION:
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 06/223,461
CC FILING DATE: 13-APR-1994
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Clark, Paul L.
CC REGISTRATION NUMBER: 301642
CC REFERENCE/JOCKET NUMBER: 00786/230001
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (617) 542-5070
CC TELEFAX: (617) 542-8906
CC TELEX: 100254
CC INFORMATION FOR SEQ ID NO: 2:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 885 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: not relevant
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC SEQUENCE 885 AA: 102110 MW: 4152599 CN:

SQ
Query Match          70.2%   Score 40: DB 4: Length 885:
Best Local Similarity 44.4%   Pred.No. 3.69e-02:
Matches           4: Conservative      Mismatches 0: Indels

Db    325 SIRRAE:I 333
QY     1 SIKRVSNI 9

```

```

Db      325 SIRRLAEI 333
      _|:::|
QY      1 SIKRVSMVI 9

```

Search completed: Tue Jun 20 13:37:51 2000  
JGD time : 5 secs.

\*\*\*\*\*  
[WATERMAN] (CM)  
\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit,  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Mon Jun 19 16:28:52 2000: MasPar time 14.36 Seconds  
Tabular output not generated. 53.146 Million cell updates/sec

Title: XUS-09-142-524A-14  
Description: (1-9) from US09142524A.pep  
Sequence: 1 SIKRVSNV: 9

Scoring table: PAM 150  
Gap 15

Searched: 721208 seqs, 100765575 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: d-pending  
1:PC1 2:U6 3:U60 4:U7 5:U83 6:U81 7:U82 8:U83 9:U84A  
10:UB4B 11:U85 12:U86 13:U87 14:U88 15:U89 16:U90 17:U91  
18:U92 19:U93 20:U94 21:U95 22:NEWP 23:NEWJ50 24:NEW08  
25:NEW09

Statistics: Mean: 17.247; Variance 40.530; scal: 0.426

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

| SUMMARIES  |       |                        |        | Description |                        | Pred. No. |  |
|------------|-------|------------------------|--------|-------------|------------------------|-----------|--|
| Result No. | Score | Query Match Length (p) | ID     | Description |                        | Pred. No. |  |
| 1          | 53    | 93.0                   | 20 10  | US-08-468-  | Sequence 35, Applicati | 4.33e-01  |  |
| 2          | 53    | 93.0                   | 20 10  | US-08-467-  | Sequence 36, Applicati | 4.33e-01  |  |
| 3          | 53    | 93.0                   | 20 4   | US-07-938-  | Sequence 35, Applicati | 4.33e-01  |  |
| 4          | 53    | 93.0                   | 20 7   | US-08-226-  | Sequence 36, Applicati | 4.33e-01  |  |
| 5          | 53    | 93.0                   | 20 10  | US-08-467-  | Sequence 36, Applicati | 4.33e-01  |  |
| 6          | 53    | 93.0                   | 20 8   | US-08-350-  | Sequence 36, Applicati | 4.33e-01  |  |
| 7          | 53    | 93.0                   | 20 10  | US-08-467-  | Sequence 36, Applicati | 4.33e-01  |  |
| 8          | 53    | 93.0                   | 31 17  | US-09-142-  | Sequence 4, Applicatio | 4.33e-01  |  |
| 9          | 53    | 93.0                   | 31 17  | US-09-142-  | Sequence 5, Applicatio | 4.33e-01  |  |
| 10         | 53    | 93.0                   | 60 4   | US-07-938-  | Sequence 62, Applicati | 4.33e-01  |  |
| 11         | 53    | 93.0                   | 60 10  | US-08-468-  | Sequence 62, Applicati | 4.33e-01  |  |
| 12         | 53    | 93.0                   | 60 7   | US-08-226-  | Sequence 62, Applicati | 4.33e-01  |  |
| 13         | 53    | 93.0                   | 60 10  | US-08-467-  | Sequence 62, Applicati | 4.33e-01  |  |
| 14         | 53    | 93.0                   | 60 10  | US-08-467-  | Sequence 62, Applicati | 4.33e-01  |  |
| 15         | 53    | 93.0                   | 60 8   | US-08-350-  | Sequence 62, Applicati | 4.33e-01  |  |
| 16         | 53    | 93.0                   | 60 10  | US-08-467-  | Sequence 62, Applicati | 4.33e-01  |  |
| 17         | 53    | 93.0                   | 80 17  | US-09-142-  | Sequence 1, Applicatio | 4.33e-01  |  |
| 18         | 53    | 93.0                   | 105 17 | US-09-142-  | Sequence 2, Applicatio | 4.33e-01  |  |
| 19         | 53    | 93.0                   | 134 17 | US-09-142-  | Sequence 3, Applicatio | 4.33e-01  |  |
| 20         | 53    | 93.0                   | 353 11 | US-08-526-  | Sequence 15, Applicati | 4.33e-01  |  |

|    |    |      |         |            |                         |          |  |
|----|----|------|---------|------------|-------------------------|----------|--|
| 21 | 53 | 93.0 | 374 10  | US-08-467- | Sequence 2, Applicatio  | 4.33e-01 |  |
| 22 | 53 | 93.0 | 374 4   | US-07-729- | Sequence 2, Applicatio  | 4.33e-01 |  |
| 23 | 53 | 93.0 | 374 10  | US-08-468- | Sequence 2, Applicatio  | 4.33e-01 |  |
| 24 | 53 | 93.0 | 374 4   | US-07-938- | Sequence 2, Applicatio  | 4.33e-01 |  |
| 25 | 53 | 93.0 | 374 10  | US-08-467- | Sequence 2, Applicatio  | 4.33e-01 |  |
| 26 | 53 | 93.0 | 374 10  | US-08-467- | Sequence 2, Applicatio  | 4.33e-01 |  |
| 27 | 53 | 93.0 | 374 8   | US-08-350- | Sequence 2, Applicatio  | 4.33e-01 |  |
| 28 | 53 | 93.0 | 374 4   | US-07-740- | Sequence 2, Applicatio  | 4.33e-01 |  |
| 29 | 53 | 93.0 | 374 7   | US-08-226- | Sequence 2, Applicatio  | 4.33e-01 |  |
| 30 | 47 | 82.5 | 555 18  | US-09-270- | Sequence 62436, Applic  | 4.33e-01 |  |
| 31 | 46 | 80.7 | 20 4    | US-07-938- | Sequence 37, Applicati  | 4.33e-02 |  |
| 32 | 46 | 80.7 | 20 10   | US-08-467- | Sequence 37, Applicati  | 4.33e-02 |  |
| 33 | 46 | 80.7 | 20 10   | US-08-467- | Sequence 37, Applicati  | 4.33e-02 |  |
| 34 | 46 | 80.7 | 20 10   | US-08-468- | Sequence 37, Applicati  | 4.33e-02 |  |
| 35 | 46 | 80.7 | 20 7    | US-08-226- | Sequence 37, Applicati  | 4.33e-02 |  |
| 36 | 46 | 80.7 | 221 18  | US-09-270- | Sequence 31937, Applic  | 4.33e-02 |  |
| 37 | 46 | 80.7 | 460 3   | US-60-142- | Sequence 1045, Applicat | 4.33e-02 |  |
| 38 | 46 | 80.7 | 813 3   | US-60-171- | Sequence 525, Applicat  | 4.33e-02 |  |
| 39 | 46 | 80.7 | 813 3   | US-60-173- | Sequence 25041, Applic  | 4.33e-02 |  |
| 40 | 46 | 80.7 | 1231 17 | US-09-150- | Sequence 19, Applicati  | 4.33e-02 |  |
| 41 | 46 | 80.7 | 1284 3  | US-60-167- | Sequence 11874, Applic  | 4.33e-02 |  |
| 42 | 46 | 80.7 | 1284 3  | US-60-171- | Sequence 997, Applicat  | 4.33e-02 |  |
| 43 | 46 | 80.7 | 1284 3  | US-60-173- | Sequence 9624, Applicat | 4.33e-02 |  |
| 44 | 46 | 80.7 | 1415 3  | US-60-153- | Sequence 1036, Applicat | 4.33e-02 |  |
| 45 | 46 | 80.7 | 1434 3  | US-60-151- | Sequence 1077, Applicat | 4.33e-02 |  |

ALIGNMENTS

RESULT 1  
ID US-08-468-940-35 STANDARD: PRT: 20 AA.

XX xxxxxx

XX

XX

DE

XX

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

Sequence 36, Application US/02468940

Sequence 36, Application US/8468440

GENERAL INFORMATION:

APPLICANT: Pollock, Joanne

APPLICANT: Bond, Julian F.

APPLICANT: Gaiman, Richard D.

APPLICANT: Kuo, Mei-Chang

APPLICANT: Yeung, Siu-wei H.

APPLICANT: Brauer, Andrew

APPLICANT: Exley, Mark A.

APPLICANT: Powers, Steven J.

TITLE OF INVENTION: Amino Acid Derivatives For

TITLE OF INVENTION: Amino Acid Derivatives For

NUMBER OF SEQUENCES: 26

CORRESPONDENCE ADDRESS:

ADDRESSEE: Immunologic Pharmaceutical Corporation, Inc.

STREET: 610 Lincoln St

CITY: Waltham

STATE: MA

COUNTRY: USA

ZIP: 02154

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/468,940

FILING DATE:

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/350,225

FILING DATE: December 5, 1994

APPLICATION NUMBER: 08/226,248

FILING DATE: April 8, 1994



CC APPLICATION NUMBER: 07/938,990  
 CC FILING DATE: September 1, 1992  
 CC APPLICATION NUMBER: PCT/US93/00139  
 CC FILING DATE: January 15, 1993  
 CC ATTORNEY/AGENT INFORMATION:  
 CC NAME: Darlene A. Vanstone  
 CC REGISTRATION NUMBER: 35,729  
 CC REFERENCE/DOCKET NUMBER: 025.6 US (IMI-028P02)  
 CC TELECOMMUNICATION INFORMATION:  
 CC TELEPHONE: (617) 466-6000  
 CC TELEFAX: (617) 466-6000  
 CC INFORMATION FOR SEQ ID NO: 36:  
 CC SEQUENCE CHARACTERISTICS:  
 CC LENGTH: 20 amino acids  
 CC TYPE: amino acid  
 CC TOPOLOGY: linear  
 CC MOLECULE TYPE: polypeptide  
 CC FRAGMENT TYPE: internal  
 CC SEQUENCE: 20 AA: 2090 MW: 2304 CN:  
 CC  
 CC Query Match 93.0% Score 53 DB 10 Length 20:  
 CC Best Local Similarity 100.0% Pred. No. 4,33e-017  
 CC Matches 8: Conservative 0: Mismatches 0: Indels 0: Gaps 0:  
 CC  
 CC Db 10 IKRVSNVI 17  
 CC III III  
 CC QY 2 IKRVSNVI 4  
 CC  
 CC RESULT 2  
 CC ID US-09-467-005-36 STANDARD: PRI: 20 AA:  
 CC XX XXXXXX  
 CC  
 CC Sequence 36, Application US/09467005  
 CC Sequence 36, Application US/09467006  
 CC GENERAL INFORMATION:  
 CC APPLICANT: Griffith, Irwin J.  
 CC APPLICANT: Pollock, Joanne  
 CC APPLICANT: Bond, Julian F.  
 CC APPLICANT: Garman, Richard D.  
 CC APPLICANT: Kuo, Mei-Chang  
 CC APPLICANT: Yeung, Su-mei H.  
 CC APPLICANT: Bauer, Andrew  
 CC APPLICANT: Exley, Mark A.  
 CC APPLICANT: Powers, Steven P.  
 CC TITLE OF INVENTION: Allergenic Proteins And Peptides From  
 CC NUMBER OF SEQUENCES: 26  
 CC CORRESPONDENCE ADDRESS:  
 CC ADDRESSEE: Immunologic Pharmaceutical Corporation, Inc.  
 CC STREET: 610 Lincoln St  
 CC CITY: Waltham  
 CC STATE: MA  
 CC COUNTRY: USA  
 CC ZIP: 02154  
 CC COMPUTER READABLE FORM:  
 CC MEDIUM TYPE: Floppy disk  
 CC COMPUTER: IBM PC compatible  
 CC OPERATING SYSTEM: PC-DOS/MS-DOS  
 CC SOFTWARE: Patent Release #1.0, Version #1.25  
 CC CURRENT APPLICATION DATA:  
 CC APPLICATION NUMBER: US/09467006  
 CC FILING DATE: June 6, 1995  
 CC CLASSIFICATION: 424  
 CC PRIOR APPLICATION DATA:  
 CC APPLICATION NUMBER: 08/350,225  
 CC FILING DATE: December 6, 1994  
 CC ATTORNEY/AGENT INFORMATION:  
 CC NAME: Jane E. Remillard

CC REGISTRATION NUMBER: 38,972  
 CC REFERENCE/DOCKET NUMBER: 025.6 USDS (IMI-028CP05)  
 CC TELECOMMUNICATION INFORMATION:  
 CC TELEPHONE: (617) 227-7400  
 CC TELEFAX: (617) 227-5941  
 CC INFORMATION FOR SEQ ID NO: 36:  
 CC SEQUENCE CHARACTERISTICS:  
 CC LENGTH: 20 amino acids  
 CC TYPE: amino acid  
 CC TOPOLOGY: linear  
 CC MOLECULE TYPE: peptide  
 CC FRAGMENT TYPE: internal  
 CC SEQUENCE: 20 AA: 2090 MW: 2304 CN:  
 CC  
 CC Query Match 93.0% Score 54 DB 10 Length 20:  
 CC Best Local Similarity 100.0% Pred. No. 4,33e-017  
 CC Matches 8: Conservative 0: Mismatches 0: Indels 0: Gaps 0:  
 CC  
 CC Db 10 IKRVSNVI 17  
 CC III III  
 CC QY 2 IKRVSNVI 9  
 CC  
 CC RESULT 3  
 CC ID US-07-938-990A-46 STANDARD: PRI: 20 AA:  
 CC XX XXXXXX  
 CC  
 CC Sequence 36, Application US/07938990A  
 CC Sequence 36, Application US/07938990A  
 CC GENERAL INFORMATION:  
 CC APPLICANT: Griffith, Irwin J.  
 CC APPLICANT: Pollock, Joanne  
 CC APPLICANT: Bond, Julian  
 CC TITLE OF INVENTION: Allergenic Proteins And Peptides From  
 CC NUMBER OF SEQUENCES:  
 CC CORRESPONDENCE ADDRESS:  
 CC ADDRESSEE: Lahive & Cockfield  
 CC STREET: Sixty State Street  
 CC CITY: Boston  
 CC STATE: MA  
 CC COUNTRY: USA  
 CC ZIP: 02109  
 CC COMPUTER READABLE FORM:  
 CC MEDIUM TYPE: Floppy disk  
 CC COMPUTER: IBM PC compatible  
 CC OPERATING SYSTEM: PC-DOS/MS-DOS  
 CC SOFTWARE: Patent Release #1.0, Version #1.25  
 CC CURRENT APPLICATION DATA:  
 CC APPLICATION NUMBER: US/07938,990A  
 CC FILING DATE: 19920901  
 CC CLASSIFICATION: 435  
 CC PRIOR APPLICATION DATA:  
 CC APPLICATION NUMBER: 07/730,452  
 CC FILING DATE: July 15, 1992  
 CC PRIOR APPLICATION DATA:  
 CC APPLICATION NUMBER: 07/729,134  
 CC FILING DATE: July 10, 1991  
 CC ATTORNEY/AGENT INFORMATION:  
 CC NAME: Amy E. Mandragoras  
 CC REGISTRATION NUMBER: 36,207  
 CC REFERENCE/DOCKET NUMBER: IPC-025CC (IMI-028)  
 CC TELECOMMUNICATION INFORMATION:  
 CC TELEPHONE: (617) 227-7400  
 CC TELEFAX: (617) 227-5941  
 CC INFORMATION FOR SEQ ID NO: 36:  
 CC SEQUENCE CHARACTERISTICS:  
 CC LENGTH: 20 amino acids  
 CC TYPE: AMINO ACID

| Query Match                  | 93.0%   | Score 53     | DB 7     | Length 20 |
|------------------------------|---|--------------|----------|-----------|
| Best Local Similarity 100.0% | Pred. No. 4,33e-01                                      |              |          |           |
| Matches 8                    | Conservative 0  | Mismatches 0 | Indels 0 | Gaps 0    |
| Db                           | 10 IKRVSNI 17   |              |          |           |
| QY                           | 2 IKRVSNI 9   |              |          |           |
| RESULT                       | 5   |              |          |           |
| ID                           | US-08-467-023-36  | STANDARD:    | PRT:     | 20 AA     |
| XX                           | xxxxxx  |              |          |           |
| AC                           | xxxxxx  |              |          |           |
| XX                           | xxxxxx  |              |          |           |
| DT                           | XX  |              |          |           |
| DE                           | Sequence 36, Application US/08457023                    |              |          |           |
| XX                           | Sequence 36, Application US/08457023                    |              |          |           |
| CC                           | GENERAL INFORMATION:                                    |              |          |           |
| CC                           | APPLICANT: Griffith, Irwin J.                           |              |          |           |
| CC                           | APPLICANT: Pollock, Joanne                              |              |          |           |
| CC                           | APPLICANT: Bond, Julian F.                              |              |          |           |
| CC                           | APPLICANT: Garman, Richard D.                           |              |          |           |
| CC                           | APPLICANT: Kuo, Mei-Chang                               |              |          |           |
| CC                           | APPLICANT: Yeung, Siu-mei H.                            |              |          |           |
| CC                           | APPLICANT: Brauer, Andrew                               |              |          |           |
| CC                           | APPLICANT: Exley, Mark A.                               |              |          |           |
| CC                           | APPLICANT: Powers, Steven P.                            |              |          |           |
| CC                           | TITLE OF INVENTION: Allergic Proteins And Peptides From |              |          |           |
| CC                           | TITLE OF INVENTION: Japanese Cedar Pollen               |              |          |           |
| CC                           | NUMBER OF SEQUENCES: 261                                |              |          |           |
| CC                           | CORRESPONDENCE ADDRESS:                                 |              |          |           |
| CC                           | ADDRESSEE: Immunologic Pharmaceutical Corporation, Inc. |              |          |           |
| CC                           | STREET: 610 Lincoln St.                                 |              |          |           |
| CC                           | CITY: Waltham   |              |          |           |
| CC                           | STATE: MA   |              |          |           |
| CC                           | COUNTRY: USA  |              |          |           |
| CC                           | ZIP: 02154  |              |          |           |
| CC                           | COMPUTER READABLE FORM:                                 |              |          |           |
| CC                           | MEDIUM TYPE: Floppy                                     |              |          |           |
| CC                           | COMPUTER: IBM PC compatible                             |              |          |           |
| CC                           | OPERATING SYSTEM: PC-DOS/MS-DOS                         |              |          |           |
| CC                           | SOFTWARE: Patent Release #1.0, Version #1.25            |              |          |           |
| CC                           | CURRENT APPLICATION DATA:                               |              |          |           |
| CC                           | APPLICATION NUMBER: US/08/467-023                       |              |          |           |
| CC                           | FILING DATE: June 6, 1995                               |              |          |           |
| CC                           | CLASSIFICATION: 424                                     |              |          |           |
| CC                           | PRIOR APPLICATION DATA:                                 |              |          |           |
| CC                           | APPLICATION NUMBER: US/08/226-248A                      |              |          |           |
| CC                           | FILING DATE: April 8, 1994                              |              |          |           |
| CC                           | CLASSIFICATION: 424                                     |              |          |           |
| CC                           | PRIOR APPLICATION DATA:                                 |              |          |           |
| CC                           | APPLICATION NUMBER: 07/936,890                          |              |          |           |
| CC                           | FILING DATE: September 1, 1992                          |              |          |           |
| CC                           | APPLICATION NUMBER: PCT/US93/00139                      |              |          |           |
| CC                           | FILING DATE: January 15, 1993                           |              |          |           |
| CC                           | ATTORNEY/AGENT INFORMATION:                             |              |          |           |
| CC                           | NAME: Darlene A. Vanstone                               |              |          |           |
| CC                           | REGISTRATION NUMBER: 35,729                             |              |          |           |
| CC                           | REFERENCE/DOCKET NUMBER: 025.5 US (IM-028CP)            |              |          |           |
| CC                           | TELECOMMUNICATION INFORMATION:                          |              |          |           |
| CC                           | TELEPHONE: (617) 466-6000                               |              |          |           |
| CC                           | TELEFAX: (617) 466-6040                                 |              |          |           |
| CC                           | INFORMATION FOR SEQ ID NO: 36:                          |              |          |           |
| CC                           | SEQUENCE CHARACTERISTICS:                               |              |          |           |
| CC                           | LENGTH: 20 amino acids                                  |              |          |           |
| CC                           | TYPE: amino acid  |              |          |           |
| CC                           | TOPOLOGY: linear  |              |          |           |
| CC                           | MOLECULE TYPE: peptide                                  |              |          |           |
| CC                           | FRAGMENT TYPE: internal                                 |              |          |           |
| CC                           | SEQUENCE 20 AA: 2080 MW: 2304 CN:                       |              |          |           |
| Query Match                  | 93.0%   | Score 53     | DB 10    | Length 20 |
| Best Local Similarity 100.0% | Pred. No. 4,33e-01                                      |              |          |           |
| Matches 8                    | Conservative 0  | Mismatches 0 | Indels 0 | Gaps 0    |
| Db                           | 10 IKRVSNI 17   |              |          |           |
| QY                           | 2 IKRVSNI 9   |              |          |           |
| RESULT                       | 5   |              |          |           |
| ID                           | US-08-467-023-36  | STANDARD:    | PRT:     | 20 AA     |
| XX                           | xxxxxx  |              |          |           |
| AC                           | xxxxxx  |              |          |           |
| XX                           | xxxxxx  |              |          |           |
| DT                           | XX  |              |          |           |
| DE                           | Sequence 36, Application US/08226248A                   |              |          |           |
| XX                           | Sequence 36, Application US/08226248A                   |              |          |           |
| CC                           | GENERAL INFORMATION:                                    |              |          |           |
| CC                           | APPLICANT: Griffith, Irwin J.                           |              |          |           |
| CC                           | APPLICANT: Pollock, Joanne                              |              |          |           |
| CC                           | APPLICANT: Bond, Julian F.                              |              |          |           |
| CC                           | APPLICANT: Garman, Richard D.                           |              |          |           |
| CC                           | APPLICANT: Kuo, Mei-Chang                               |              |          |           |
| CC                           | APPLICANT: Yeung, Siu-mei H.                            |              |          |           |
| CC                           | APPLICANT: Brauer, Andrew                               |              |          |           |
| CC                           | APPLICANT: Exley, Mark A.                               |              |          |           |
| CC                           | APPLICANT: Powers, Steven P.                            |              |          |           |
| CC                           | TITLE OF INVENTION: Allergic Proteins And Peptides From |              |          |           |
| CC                           | TITLE OF INVENTION: Japanese Cedar Pollen               |              |          |           |
| CC                           | NUMBER OF SE  |              |          |           |

```

17 2 KRVSNI 9
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040
1041
1042
1043
1044
1045
1046
10
```

```

RESULT 7
XX US-08-467-697-36 STANDARD: PRT: 20 AA.
XX
XX
XX
XX
XX
XX
XX
XX
XX
XX
Sequence 36, Application US/8467697
Sequence 36, Application US/8467697
GENERAL INFORMATION:
CC APPLICANT: Griffith, Lisa J.
CC APPLICANT: Pollock, Joan
CC APPLICANT: Bond, Julian
CC APPLICANT: Garton, Richard
CC APPLICANT: Kuo, Mei-Cha
CC APPLICANT: Yeung, Siu-mu
CC APPLICANT: Brauer, Andre
CC APPLICANT: Exley, Mark A.
CC APPLICANT: Powers, Steven V.
CC TITLE OF INVENTION: Allergic Proteins And Peptides From
CC TITLE OF INVENTION: Japanese Cedar Pollen
CC NUMBER OF SEQUENCES: 251
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Immunologic Pharmaceutical Corporation, Inc.
CC STREET: 610 Lincoln St
CC CITY: Waltham
CC STATE: MA
CC COUNTRY: USA
CC ZIP: 02154
CC
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patent In Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/467,697
CC FILING DATE: June 6, 1995
CC CLASSIFICATION: 424
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 387,60,225
CC FILING DATE: December 6, 1994
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Jane E. Remilla
CC REGISTRATION NUMBER: 872
CC REFERENCE/DOCKET NUMBER: 025.6 USN (LMI-02807P14)
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (617) 227-6341
CC TELEFAX: (617) 227-6341
CC INFORMATION FOR SEQ ID NO: 36:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 20 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC FRAGMENT TYPE: internal
CC
CC
CC
CC
CC
CC
CC
CC
CC
CC
Sequence 20 AA: 2080 MW: 2304 CN:
Query Match 93.0%; Score 53; DB 10; Length 20;
Best Local Similarity 100.0%; Pred. No. 4.31e+01;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 10 IKRVSNVI 17
QY 2 IKRVSNV; 9
XXXXXXXX
RESULT 8
XX US-09-142-524-4 STANDARD: PRT: 31 AA.
XX
XX
XX
XX

```

```

XX  Query Match          93.0%: Score 53: DB 17: Length 31:
DE  Best Local Similarity 100.0%: Pred. No. 4.33e+01:
XX  Matches 8: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

CC  Sequence 4, Application US/09142524
CC  GENERAL INFORMATION:
CC  APPLICANT: Sore, Toshio
CC  APPLICANT: Kume, Akio
CC  APPLICANT: Kairiki, Kazuo
CC  APPLICANT: Iwama, Akiko
CC  APPLICANT: Kico, Kohsuke
CC  TITLE OF INVENTION: Peptide-based Immunotherapeutic Agent For Treating
CC  TITLE OF INVENTION: Allergic Diseases
CC  FILE REFERENCE: Docket No. SPO-103
CC  CURRENT APPLICATION NUMBER: US/09/142.524
CC  CURRENT FILING DATE: 1999-01-04
CC  EARLIER APPLICATION NUMBER: 8/80/702
CC  EARLIER FILING DATE: 1996-03-10
CC  EARLIER APPLICATION NUMBER: PCT/JP97/00740
CC  EARLIER FILING DATE: 1997-03-10
CC  NUMBER OF SEQ ID NOS: 5
CC  SOFTWARE: PatentIn Ver. 2.0
CC  SEQ ID NO 4
CC  LENGTH: 31
CC  TYPE: PRI
CC  ORGANISM: Unknown
CC  FEATURE:
CC  OTHER INFORMATION: Description of Unknown Organism:peptide
CC  SEQUENCE 3: AA: 3661 MW: 4960 CN:

XX  Query Match          93.0%: Score 53: DB 17: Length 31:
DE  Best Local Similarity 100.0%: Pred. No. 4.33e+01:
XX  Matches 8: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

DB  24 IKRVSNV: 31
QY  2 IKRVSNV: 9

RESULT 9
ID  US-09-142-524-5 STANDARD: PRI: 3: AA:
XX  Query Match          93.0%: Score 53: DB 17: Length 31:
XX  Best Local Similarity 100.0%: Pred. No. 4.33e+01:
XX  Matches 8: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

DE  24 IKRVSNV: 31
QY  2 IKRVSNV: 9

Sequence 5, Application US/09142524
GENERAL INFORMATION:
APPLICANT: Sore, Toshio
APPLICANT: Kume, Akio
APPLICANT: Kairiki, Kazuo
APPLICANT: Iwama, Akiko
APPLICANT: Kico, Kohsuke
TITLE OF INVENTION: Peptide-based Immunotherapeutic Agent For Treating
TITLE OF INVENTION: Allergic Diseases
FILE REFERENCE: Docket No. SPO-103
CURRENT APPLICATION NUMBER: US/09/142.524
CURRENT FILING DATE: 1999-01-04
EARLIER APPLICATION NUMBER: 8/80/702
EARLIER FILING DATE: 1996-03-10
EARLIER APPLICATION NUMBER: PCT/JP97/00740
EARLIER FILING DATE: 1997-03-10
NUMBER OF SEQ ID NOS: 5
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 5
LENGTH: 31
TYPE: PRI
ORGANISM: Unknown
FEATURE:
OTHER INFORMATION: Description of Unknown Organism:peptide

```

```

S0  SEQUENCE 3: AA: 3585 MW: 4902 CN:

Query Match          93.0%: Score 53: DB 17: Length 31:
Best Local Similarity 100.0%: Pred. No. 4.33e+01:
Matches 8: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

DB  24 IKRVSNV: 31
QY  2 IKRVSNV: 9

RESULT 10
ID  US-07-938-990A-62 STANDARD: PRI: 60 AA:
XX  Query Match          93.0%: Score 53: DB 4: Length 60:
XX  Best Local Similarity 100.0%: Pred. No. 4.33e+01:
XX  Matches 8: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

DE  Sequence 62, Application US/07938990A
CC  GENERAL INFORMATION:
CC  APPLICANT: Griffith, Irwin J.
CC  APPLICANT: Pollock, Joanne
CC  APPLICANT: Bond Julian
CC  TITLE OF INVENTION: Allergenic Proteins And Peptides From
CC  TITLE OF INVENTION: Japanese Cedar Pollen
CC  NUMBER OF SEQUENCES: 70
CC  CORRESPONDENCE ADDRESS:
CC  ADDRESSEE: Lahive & Cockfield
CC  STREET: Sixty State Street
CC  CITY: Boston
CC  STATE: MA
CC  COUNTRY: USA
CC  ZIP: 02109
CC  COMPUTER READABLE FORM:
CC  MEDIUM TYPE: Floppy disk
CC  COMPUTER: IBM PC compatible
CC  OPERATING SYSTEM: PC-DOS/MS-DOS
CC  SOFTWARE: PatentIn Release #1.0, Version #1.25
CC  CURRENT APPLICATION DATA:
CC  APPLICATION NUMBER: US/07/938.990A
CC  FILING DATE: 19920901
CC  CLASSIFICATION: 435
CC  PRIOR APPLICATION DATA:
CC  PRIOR APPLICATION NUMBER: 07/374,464
CC  FILING DATE: July 16, 1992
CC  PRIOR APPLICATION DATA:
CC  APPLICATION NUMBER: 07/374,464
CC  FILING DATE: July 16, 1992
CC  ATTORNEY/AGENT INFORMATION:
CC  NAME: Amy E. Mandragoras
CC  REGISTRATION NUMBER: 36,207
CC  REFERENCE/DOCKET NUMBER: IPC-025CC (IMI-028)
CC  TELEPHONE: (617) 227-5941
CC  TELEFAX: (617) 227-7400
CC  INFORMATION FOR SEQ ID NO: 62:
CC  SEQUENCE CHARACTERISTICS:
CC  LENGTH: 60 amino acids
CC  TYPE: AMINO ACID
CC  TOPOLOGY: linear
CC  MOLECULE TYPE: peptide
CC  FRAGMENT TYPE: internal
CC  SEQUENCE 50 AA: 5644 MW: 19464 CN:

S0  SEQUENCE 50 AA: 5644 MW: 19464 CN:

Query Match          93.0%: Score 53: DB 4: Length 60:
Best Local Similarity 100.0%: Pred. No. 4.33e+01:
Matches 8: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

DB  50 IKRVSNV: 57
QY  2 IKRVSNV: 9

```

```

RESULT 11
ID US-09-458-940-62 STANDARD: PRT: 60 AA
XX
AC xxxxxx
XX
DT
XX
XX
DE Sequence 62, Application US/08466940
XX
CC Sequence 62, Application US/08466940
CC GENERAL INFORMATION:
CC APPLICANT: Griffith, Irwin J.
CC APPLICANT: Pollock, Joanne
CC APPLICANT: Bond, Julian F.
CC APPLICANT: Garman, Richard O.
CC APPLICANT: Kuo, Mei-Chang
CC APPLICANT: Yeung, Siu-mei H.
CC APPLICANT: Brauer, Andrew
CC APPLICANT: Exley, Mark A.
CC APPLICANT: Powers, Steven P.
CC TITLE OF INVENTION: Allergenic Proteins And Peptides From
CC TITLE OF INVENTION: Japanese Cedar Pollen
CC NUMBER OF SEQUENCES: 26
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Immunologic Pharmaceutical Corporation, Inc.
CC STREET: 610 Lincoln St
CC CITY: Waltham
CC STATE: MA
CC COUNTRY: USA
CC ZIP: 02154
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/05/458,940
CC FILING DATE:
CC CLASSIFICATION: 424
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/350,225
CC FILING DATE: December 6, 1994
CC APPLICATION NUMBER: 08/226,248
CC FILING DATE: April 9, 1994
CC APPLICATION NUMBER: 07/638,940
CC FILING DATE: September 1, 1992
CC APPLICATION NUMBER: PCT/US93/00179
CC FILING DATE: January 15, 1993
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Darlene A. Vansuto
CC REGISTRATION NUMBER: 35,759
CC REFERENCE/CKET NUMBER: 025.6 US (IMI-024CP2)
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (617) 466-6000
CC TELEFAX: (617) 466-6040
CC INFORMATION FOR SEQ ID NO: 62:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 60 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC FRAGMENT TYPE: internal
CC SEQUENCE 60 AA: 6644 MW: 19454 CN:

Query Match 93.0% Score 53: DB 10: Length 60:
Best Local Similarity 100.0% Pred. No. 4,33e-01:
Matches 8: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

Db 50 IKRVSNI: 57
IIIIIII
QY 2 IKRVSNI: 9

```

```

RESULT 12
ID US-08-226-248A-62 STANDARD: PRT: 60 AA
XX
AC xxxxxx
XX
DT
XX
XX
DE Sequence 62, Application US/08226248A
XX
CC Sequence 62, Application US/08226248A
CC GENERAL INFORMATION:
CC APPLICANT: Griffith, Irwin J.
CC APPLICANT: Pollock, Joanne
CC APPLICANT: Bond, Julian F.
CC APPLICANT: Garman, Richard O.
CC APPLICANT: Kuo, Mei-Chang
CC APPLICANT: Yeung, Siu-mei H.
CC APPLICANT: Brauer, Andrew
CC APPLICANT: Exley, Mark A.
CC APPLICANT: Powers, Steven P.
CC TITLE OF INVENTION: Allergenic Proteins And Peptides From
CC TITLE OF INVENTION: Japanese Cedar Pollen
CC NUMBER OF SEQUENCES: 26
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Immunologic Pharmaceutical Corporation, Inc.
CC STREET: 610 Lincoln St
CC CITY: Waltham
CC STATE: MA
CC COUNTRY: USA
CC ZIP: 02154
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: 08/226,248A
CC FILING DATE: April 9, 1994
CC CLASSIFICATION: 424
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 07/638,940
CC FILING DATE: September 1, 1992
CC APPLICATION NUMBER: PCT/US93/00179
CC FILING DATE: January 15, 1993
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Darlene A. Vansuto
CC REGISTRATION NUMBER: 35,759
CC REFERENCE/CKET NUMBER: 025.6 US (IMI-024CP2)
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (617) 466-6000
CC TELEFAX: (617) 466-6040
CC INFORMATION FOR SEQ ID NO: 62:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 60 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC FRAGMENT TYPE: internal
CC SEQUENCE 60 AA: 6644 MW: 19454 CN:

Query Match 93.0% Score 53: DB 7: Length 60:
Best Local Similarity 100.0% Pred. No. 4,33e-01:
Matches 8: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

Db 50 IKRVSNI: 57
IIIIIII
QY 2 IKRVSNI: 9

RESULT 13
ID US-08-467-697-62 STANDARD: PRT: 60 AA

```



CC APPLICANT: Garman, Richard D:  
 CC APPLICANT: Kuo, Mei-Chang:  
 CC APPLICANT: Yeung, Siu-mei H.:  
 CC APPLICANT: Brauer, Andrew:  
 CC APPLICANT: Exley, Mark A.:  
 CC APPLICANT: Powers, Steven P.  
 CC TITLE OF INVENTION: Allergenic Proteins And Peptides From  
 CC TITLE OF INVENTION: Japanese Cedar Pollen  
 CC NUMBER OF SEQUENCES: 261  
 CC CORRESPONDENCE ADDRESS:  
 CC ADDRESSEE: Immunologic Pharmaceutical Corporation, Inc.  
 CC STREET: 610 Lincoln St  
 CC CITY: Wallingham  
 CC STATE: MA  
 CC COUNTRY: USA  
 CC ZIP: 02154  
 CC COMPUTER READABLE FORM:  
 CC MEDIUM TYPE: Floppy disk  
 CC COMPUTER: IBM PC Compatible  
 CC OPERATING SYSTEM: PC-DOS/MS-DOS  
 CC SOFTWARE: PatentIn Release #1.0, Version #1.25  
 CC CURRENT APPLICATION DATA:  
 CC APPLICATION NUMBER: US/08/350,225  
 CC FILING DATE: December 6, 1994  
 CC CLASSIFICATION: 424  
 CC PRIOR APPLICATION DATA:  
 CC APPLICATION NUMBER: 08/226,246  
 CC FILING DATE: April 8, 1994  
 CC APPLICATION NUMBER: 07/918,990  
 CC FILING DATE: September 1, 1992  
 CC APPLICATION NUMBER: PCI/US93/00139  
 CC FILING DATE: January 25, 1993  
 CC ATTORNEY/AGENT INFORMATION:  
 CC NAME: Darlene A. VanStoore  
 CC REGISTRATION NUMBER: 35,729  
 CC REFERENCE/DOCKET NUMBER: 025.6 US (IMI-028CPZ)  
 CC TELECOMMUNICATION INFORMATION:  
 CC TELEPHONE: (617) 466-6000  
 CC TELEFAX: (617) 466-6040  
 CC INFORMATION FOR SEQ ID NO: 62:  
 CC SEQUENCE CHARACTERISTICS:  
 CC LENGTH: 60 amino acids  
 CC TYPE: amino acid  
 CC TOPOLOGY: linear  
 CC MOLECULE TYPE: peptide  
 CC FRAGMENT TYPE: internal  
 CC SEQUENCE: 60 AA: 6644 MW: 19464 DA:

Query Match 94.0% Score 531 DB 81 Length 600  
 Best Local Similarity 100.0% Prod. No. 4-940-01  
 Matches 61 Conservations 0 Mismatches 0 Indels 0 Gaps 0

Db 50 IKRVSNI 57  
 1111111  
 QY 2 IKRVSNI 9

Search completed: Mon Jun 19 16:29:08 2000  
 Job time: 16 secs.

```
*****
W A S E L L
(1M)
*****

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (C) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

vsearch_pf protein - protein database search, using Smith-Waterman algorithm

Run on: Mon Jun 19 16:16:47 2000: Maspar time 4.95 Seconds
Regular output not generated. 85.798 Million cell updates/sec
```

|    |    |      |      |   |        |                         |          |
|----|----|------|------|---|--------|-------------------------|----------|
| 24 | 43 | 75.4 | 318  | 2 | S37225 | ribose-phosphate pyro   | 9.57e-01 |
| 25 | 43 | 75.4 | 377  | 2 | G64937 | lipopolysaccharide bi   | 9.57e-01 |
| 26 | 43 | 75.4 | 390  | 2 | S.87   | hypothetical protein    | 9.57e-01 |
| 27 | 43 | 75.4 | 431  | 2 | G.24   | probable aspartokinase  | 9.57e-01 |
| 28 | 43 | 75.4 | 459  | 2 | I.108  | hypothetical protein    | 9.57e-01 |
| 29 | 43 | 75.4 | 583  | 1 | I.14   | hypothetical protein    | 9.57e-01 |
| 30 | 43 | 75.4 | 784  | 1 | A1704  | endopeptidase La 15C    | 9.57e-01 |
| 31 | 43 | 75.4 | 785  | 1 | Q6V28  | DS protein - vaccinia   | 9.57e-01 |
| 32 | 43 | 75.4 | 785  | 2 | B1347  | ATP/GTP-binding prote   | 9.57e-01 |
| 33 | 43 | 75.4 | 785  | 2 | D4255  | GTP protein - vaccini   | 9.57e-01 |
| 34 | 43 | 75.4 | 792  | 2 | F06487 | potassium transport p   | 9.57e-01 |
| 35 | 42 | 73.7 | 256  | 2 | A2127  | beta-lactamase (EC 3.   | 9.57e-01 |
| 36 | 42 | 73.7 | 269  | 2 | A2129  | hypothetical protein    | 9.57e-01 |
| 37 | 42 | 73.7 | 313  | 2 | A1284  | lysophospholipase L2    | 9.57e-01 |
| 38 | 42 | 73.7 | 313  | 1 | A2137  | hypothetical protein    | 9.57e-01 |
| 39 | 42 | 73.7 | 366  | 2 | G.157  | hypothetical protein    | 9.57e-01 |
| 40 | 42 | 73.7 | 433  | 2 | S.31   | keratin, type I, epide  | 9.57e-01 |
| 41 | 42 | 73.7 | 503  | 2 | H.33   | [L-xylose transport] A: | 9.57e-01 |
| 42 | 42 | 73.7 | 954  | 2 | I.37   | glycoprotein B - alic   | 9.57e-01 |
| 43 | 42 | 73.7 | 1054 | 2 | S.29   | probable isoleucyl-tR   | 9.57e-01 |
| 44 | 42 | 73.7 | 1177 | 2 | P.130  | chromosome segregatio   | 9.57e-01 |
| 45 | 42 | 73.7 | 1179 | 2 | P.140  | probable chromosome     | 9.57e-01 |

## ACKNOWLEDGMENTS

```

RESULT      1
ENTRY       JC2124      #type complete
ENTRY       Major allergen Cry j I precursor (clone pCC1-13) - Japanese
ENTRY       cedar
ENTRY       #formal_name Cryptomeria japonica #common_name Japanese cedar
ENTRY       14-Jul-1994 #sequence_revision 14-Jul-1994 #text_change
ENTRY       26-Aug-1999
ACCESSIONS JC2124
REFERENCE   JC2123
ENTRY       Sone, T.; Komiyaama, N.; Shimizu, K.; Kusakabe, T.; Morikubo,
ENTRY       K.; Kiro, K.
ENTRY       #journal Biochem. Biophys. Res. Commun. (1994) 199:619-625
ENTRY       #title Cloning and sequencing of cDNA coding for Cry j I, a major
ENTRY       allergen of Japanese cedar pollen.
ENTRY       #cross-references NCID:94183234

```



```

ACCESSIONS      25-Aug-1994
REFERENCE       J02123; PC2056
AUTHORS        Sano, T.; Komiyama, N.; Shimizu, K.; Kusakabe, T.; Morikubo,
                K.; Kido, K.
JOURNAL        Biochem. Biophys. Res. Commun. (1994) 199:619-625
TITLE          Cloning and sequencing of cDNA coding for Cry j 1, a major
                allergen of Japanese cedar pollen.
CROSS-REFERENCES MIM:9418234
ACCESSION      J02123
**molecule_type** RNA
**residues**      1-374 **label** SON
**cross-references** GB:025544; NID:9494531; PID:0106086; PID:9494632
**experimental_source** pollen
ACCESSION      PC2765
**molecule_type** protein
**residues**      22-5358 **label** 299-337:345-372 **label** S02
**note**          the authors described carbohydrate binding site for
                residue 279
CLASSIFICATION  as per family pectate lyase LA:59
KEYWORDS        glycoprotein; pollen
FEATURE
1-21            *domain: signal sequence *status predicted *label: SIGN
22-374          *product: major allergen Cry j 1 (clone PCOL-2-2) *status
                predicted *label: MA1N
158-191:293-354 *binding site: carbohydrate (Asn) (covalent) *status
                predicted
SUMMARY         *length 374 *molecular_weight 40645 *checksum 2920
Query Match    93.0% Score 53; DB 2; Length 374;
Best Local Similarity 100.0%; Pred. No. 3,776-00;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 131 IKRVSNVI 148
QY 2 IKRVSNVI 9
RESULT 2
ENTRY      S01509 *type complete
TITLE      Pectate lyase - Aspergillus sp.
ORGANISM   Aspergillus sp.
DATE       15-Jul-1995 *sequence_revision 19-Apr-1996 *text_change
19-Apr-1996
ACCESSIONS S01509
AUTHORS     Hol, M.C.; Whitehead, M.P.; Clevehead, T.E.; Dean, M.A.;
                Goff, Genet. (1995) 27:142-149
JOURNAL     Sequence analysis of the Aspergillus nidulans pectate lyase
                cDNA gene and evidence for binding of pectate lyase to
                pectin. A recombinant of curban carbohydrate repression.
CROSS-REFERENCES MIM:9494536
ACCESSION      S01509 pre-initiary
**status**
**molecule_type** DNA
**residues**      1-326 **label** HIM
SUMMARY       *length 326 *molecular_weight 34565 *checksum 3698
Query Match    87.7% Score 50; DB 2; Length 326;
Best Local Similarity 87.5%; Pred. No. 1,856-00;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Db 111 IKRVSNVI 118
QY 2 IKRVSNVI 9
RESULT 4
ENTRY      S73776 *type complete
TITLE      NC269 homolog Fil-Orf358b - Mycoplasma pneumoniae (ATCC
                29342) (S0C3)
ORGANISM     Mycoplasma pneumoniae
VARIETY      ATCC 29342

```

```

DATE          27-Feb-1997 *sequence_revision 25-Apr-1997 *text_change
17-Jul-1998
ACCESSIONS    S73776
REFERENCE      S73327
AUTHORS        Himmelreich, R.; Hilbert, H.; Plagens, H.; Fritsch, E.;
                B.C.; Herrmann, K.
JOURNAL        Nucleic Acids Res. (1996) 24:4420-4449
TITLE          Complete sequence analysis of the genome of the bacterium
                Mycoplasma pneumoniae.
CROSS-REFERENCES MIM:019705885
ACCESSION      S73776
**status**      preliminary; nucleic acid sequence not shown;
                translation not shown
**molecule_type** DNA
**residues**      1-358 **label** HIM
**cross-references** EMBL:AE000441; GB:000089; NID:01674140; PID:01674148
**note**          the nucleotide sequence was submitted to the EMBL data
                library, November 1996
GENETICS
**genetic_code** S0C3
SUMMARY        *length 358 *molecular_weight 42617 *checksum 8572
Query Match    82.5% Score 47; DB 2; Length 358;
Best Local Similarity 77.8%; Pred. No. 8,568-00;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 349 SIKRVSKL 357
QY 1 SIKRVSNVI 9
RESULT 5
ENTRY      E70414 *type complete
TITLE      hypothetical protein aq_1324 - Aquifex aeolicus
ORGANISM     Aquifex aeolicus
DATE         08-May-1998 *seq. CE_revision 08-May-1998 *text_change
08-May-1998
ACCESSIONS    E70414
REFERENCE      Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.;
                Lenox, A.; Graham, D.E.; Overbeck, R.; Speed, M.A.;
                Keller, M.; Anjard, M.; Huter, R.; Feldman, R.A.; Stahl,
                J.M.; Olsen, S.; Scauston, P.V.
JOURNAL        Nature (1998) 392:153-158
TITLE          The complete genome of the hyperthermophilic bacterium
                Aquifex aeolicus
CROSS-REFERENCES MIM:028196628
ACCESSION      E70414
**status**      preliminary; nucleic acid sequence not shown;
                translation not shown
**molecule_type** DNA
**residues**      1-352 **label** A1
**cross-references** GB:AE000341; MIM:01674141; PID:02819745; GB:AE000657
**experimental_source** strain VHS
GENETICS
**gene**        aq_1324
SUMMARY        *length 192 *molecular_weight 22076 *checksum 754
Query Match    78.9% Score 45; DB 2; Length 192;
Best Local Similarity 66.7%; Pred. No. 2,326-01;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
Db 111 AKRVKNVL 119
QY 1 SIKRVSNVI 9
RESULT 6
ENTRY      D69374 *type complete
TITLE      type II secretion system protein (gspE-3) homolog -
                Archaeoglobus fulgidus
ORGANISM     Archaeoglobus fulgidus
VARIETY      *formal name Archaeoglobus fulgidus
DATE         05-Dec-1997 *sequence_revision 05-Dec-1997 *text_change

```

29-Sep-1999

ACCESSIONS  
REFERENCE  
#authors

D59374  
A59250  
Klenk, H.P.; Clayton, R.A.; Omb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson, R.J.; Gwinn, M.; Hickey, E.K.; Peterson, J.D.; Richardson, D.L.; Kierlavage, A.R.; Graham, D.E.; Kyriakides, N.C.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.; Dougherty, B.A.; McKenry, K.; Adams, M.D.; Loftus, B.; Peterson, S.; Reich, C.I.; McNeil, L.K.; Badger, J.H.; Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.; Utterback, T.; Cotton, M.D.; Spriggs, T.; Artach, P.; Kaine, B.P.; Sykes, S.M.; Sadow, P.W.; D'Andrea, K.P.; Bowman, C.; Fujii, C.; Garland, S.A.; Mason, T.M.; Olsen, G.J.; Fraser, C.M.; Smith, H.O.; Woese, C.R.; Venter, J.C.

#journal Nature (1997) 390:364-370  
#title The complete genome sequence of the hyperthermophilic sulfate-reducing archaeon *Archaeoglobus fulgidus*.

#cross-references MJ01:98049343  
#accession D59374  
#status preliminary; nucleic acid sequence not shown; translation not shown

#molecule\_type DNA  
#residues 1-569 #label KLE  
#cross-references GB:AE001035; GB:AE000782; NID:q2689358; PIDN:AB90245.1; PID:q2549599; T:GR:AF0996  
CLASSIFICATION #superfamily conserved hypothetical protein MJ0900  
SUMMARY #length 569 #molecular\_weight 56520 #checksum 121

Query Match 78.9% Score 45; DB 2; Length 569;  
Best Local Similarity 75.0% Pred. No. 2,328+01;  
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 544 EKRVSNI: 551  
Qy 2 IKRVSNVI 9

RESULT 7

ENTRY #70437 #type complete  
TITLE nucleoside-diphosphate kinase (EC 2.7.4.6) - Aquifex aeolicus  
ORGANISM #formal\_name Aquifex aeolicus  
DATE 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change

ACCESSIONS F70437  
REFERENCE A70100  
#authors Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Overbeek, R.; Sneed, M.A.; Keller, M.; Aubay, M.; Huber, R.; Feldman, R.A.; Short, J.M.; Olsen, G.J.; Swanson, R.V.

#journal Nature (1998) 392:353-359  
#title The complete genome of the hyperthermophilic bacterium *Aquifex aeolicus*.

#cross-references MJ01:98146566  
#accession F70437  
#status preliminary; nucleic acid sequence not shown; translation not shown

#molecule\_type DNA  
#residues 1-142 #label AOF  
#cross-references GB:AE000746; NID:q2983925; PIDN:AC07481.1; PID:q2983932; GB:AE0000657  
#experimental\_source strain VF5

GENETICS

#gene ndk  
CLASSIFICATION #superfamily nucleoside-diphosphate kinase  
KEYWORDS A.P. binding; hexamer; phosphatidyl; phosphoprotein; phosphotransferase; pyrimidine nucleotide biosynthesis

FEATURE

14-18 #region ATP binding #status predicted  
120 #active\_site His (phosphohistidine intermediate) #status predicted  
SUMMARY #length 142 #molecular\_weight 15942 #checksum 8862

Query Match 77.2% Score 44; DB 1; Length 142;  
Best Local Similarity 55.6% Pred. No. 3,75e+01;  
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 82 AKRVREII 90  
Qy 1 SIKRVSNVI 9

RESULT 8

ENTRY R3KT5 #type complete  
TITLE ribosomal protein S5 - *Cyanophora paradoxa* cyanobelle  
ORGANISM #formal\_name cyanobelle *Cyanophora paradoxa*  
DATE 31-Dec-1991 #sequence\_revision 31-Dec-1991 #text\_change

ACCESSIONS S12220; T06876

REFERENCE S12211  
#authors Michalowski, C.H.; Pienzad, B.; Loeffelhardt, W.; Bohmert, H.J.

#journal Mol. Gen. Genet. (1990) 224:222-231  
#title The cyanobelle S10 spc ribosomal protein gene cluster from *Cyanophora paradoxa*.

#cross-references MJ01:9117189

#accession S12220

#molecule\_type DNA

#residues 1-169 #label MIC

#cross-references GB:M30487; NID:q336645; PIDN:AAA61429.1; PID:q336655  
Z15840

#authors Stirewalt, V.L.; Michalowski, C.B.; Luffelhardt, W.; Bohmert, H.J.; Bryant, D.A.

#submission submitted to the EMBL Data Library, July 1995

#description Nucleotide sequence of the cyanobelle genome from *Cyanophora paradoxa*.

#accession T06876

#status preliminary; translated from GB/EMBL/DDBJ

#molecule\_type DNA

#residues 1-169 #label STI

#cross-references EMBL:030131; NID:q1014083; PIDN:AAA61219.1; PID:q101432

#experimental\_source strain: Pfingsheim LB555

GENETICS

#gene rps5

#genome cyanobelle

CLASSIFICATION #superfamily Escherichia coli ribosomal protein S5

KEYWORDS cyanobelle; protein biosynthesis; ribosome

SUMMARY #length 169 #molecular\_weight 17980 #checksum 7641

Query Match 77.2% Score 44; DB 1; Length 169;  
Best Local Similarity 62.5% Pred. No. 3,75e+01;  
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 24 IKRVSKVV 31  
Qy 2 IKRVSNVI 9

RESULT 9

ENTRY S77486 #type complete  
TITLE ribosomal protein S5 - *Synechocystis* sp. (strain PCC 5803)  
ORGANISM #formal\_name Synechocystis sp.  
#variety PCC 5803

DATE 25-Apr-1997 #seq ence\_revision 25-Apr-1997 #text\_change  
13-Aug-1999

ACCESSIONS S77486

REFERENCE S74322

#authors Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asai, Y.;

Nakamura, Y.; Miyajima, N.; Hikosawa, M.; Sugita, M.;

Sasanoto, S.; Kimura, T.; Hosouchi, T.; Matsuo, A.;

Muraki, A.; Nakazaki, N.; Naruo, K.; Okumura, S.; Shino, S.;

Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.;

Yasuda, M.; Tabata, S.

```

#journal: DNA Res. (1996) 3:109-136
#title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis sp. PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions.
#cross-references: NCBI:97061201
#accession: S77485
#status: preliminary; nucleic acid sequence not shown; translation not shown
#molecule_type: DNA
#residues: 1-173 #label: KAN
#cross-references: EMBL:D92905; GB:AB001339; NID:q1652360; PDB:1A018066; PDB:q1652411
#note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

GENETICS
#gene: rps5
#classification: #superfamily Escherichia coli ribosomal protein S5
#keywords: Protein biosynthesis; ribosome
#summary: #length: 173 #molecular-weight: 16241 #checksum: 674

Query Match 77.2% Score 44; DB 2; Length 173;
Best Local Similarity 62.5% Pred. No. 3.76e+01;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 24 SKRISVNV 41
QY 2 SKRISVNV 9

RESULT 10
ENTRY 101215 #type: complete
TITLE Cyclic GMP-dependent protein kinase homolog F6N23.2 - Arabidopsis thaliana
#formal_name: Arabidopsis thaliana #common_name: mouse-ear cress
#map_position: 12-Feb-1999 #sequence_revision: 12-Feb-1999 #text_change: 20-Sep-1999
ACCESSIONS 101215
REFERENCE 101215
#authors: Jaisel, C.
#submission: Submitted to the EMBL Data Library, April 1998
#description: The sequence of A. thaliana F6N23.
#accession: 101215
#status: preliminary; translated from GB/ YRI/UDBJ
#molecule_type: DNA
#residues: type 1-175 #label: GSI
#cross-references: EMBL:AF05919; NID:q3047100; PDB:q3047101; GSPS1GN0003; AISD:F6N23.2
GENETICS
#gene: AISD:F6N23.2
#introns: 6/13 37/13 67/23 87/13 104/3
#summary: #length: 175 #molecular-weight: 16241 #checksum: 3611

Query Match 77.2% Score 44; DB 2; Length 175;
Best Local Similarity 55.6% Pred. No. 3.76e+01;
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 2 SKRISVNV 31
QY 1 SKRISVNV 9

RESULT 11
ENTRY A64377 #type: complete
TITLE fumarate hydratase (EC 4.2.1.2) - Methanococcus jannaschii
#formal_name: Methanococcus jannaschii
#submission: 13-Sep-1996 #sequence_revision: 13-Sep-1996 #text_change: 16-Jul-1999
#accession: A64377
#reference: A64300
#authors: Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake, J.A.; FitzGerald, L.M.; Clayton,

```

```

R.A.; Gocayne, J.D.; Kerlavage, A.R.; Dougherty, P.A.; Tomb, J.F.; Adams, M.D.; Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weissstock, K.G.; Merrick, J.M.; Binkley, A.; Scott, J.L.; Geohagen, N.S.M.; Weidman, J.F.; Finkelman, J.L.; Nguyen, D.; Utterback, J.R.; Kelley, J.M.; Petersen, J.B.; Sadow, P.W.; Hanna, M.G.; Cotton, M.G.; Klotzel, K.M.; Hurst, M.A.; Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, R.O.; Woese, C.R.; Venturi, J.C. Science (1996) 273:1068-1073
#journal: Complete genome sequence of the methanogenic archaeon Methanococcus jannaschii.
#accession: A64377
#status: preliminary; nucleic acid sequence not shown; translation not shown
#molecule_type: DNA
#residues: 1-195 #label: BUL
#cross-references: GB:U67520; CR:17717; NID:q1591325; PDB:AA04612.1; PDB:q1591328; TIGR:MJC017; PDB:q151699

GENETICS
#map_position: F0R547001-547588
#start_codon: TIG
#classification: #superfamily iron-dependent tartrate dehydratase beta chain; iron-dependent tartrate dehydratase beta chain homology; carbon-oxygen lyase; hydrolase
#keywords: #domain iron-dependent tartrate dehydratase beta chain homology; #label: TIGP
#summary: #length: 195 #molecular-weight: 21866 #checksum: 6962

Query Match 77.2% Score 44; DB 2; Length 195;
Best Local Similarity 75.0% Pred. No. 3.76e+01;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 136 SVKRYDNU 143
QY 1 SKRISVNV 8

RESULT 12
ENTRY E70430 #type: complete
TITLE Hypothetical protein aq1501 Aquifex aeolicus
#formal_name: Aquifex aeolicus
#map_position: 06-May-1998 #sequence_revision: 06-May-1998 #text_change: 06-May-1998
ACCESSIONS E70430
REFERENCE A70300
#authors: Deckert, G.; Walling, P.V.; Gaasterland, T.; Young, W.G.; Jorck, A.; Chisholm, D.; Overbeek, R.; Shedd, M.A.; Kollman, M.A.; Bork, P.; Bult, C.J.; Fleischmann, R.D.; White, O.; Olsen, G.J.; Zhou, L.; Venturi, J.C. Science (1996) 273:1068-1073
#journal: Complete genome sequence of the hyperthermophilic bacterium Aquifex aeolicus
#accession: E70430
#status: preliminary; nucleic acid sequence not shown; translation not shown
#molecule_type: DNA
#residues: 1-256 #label: AUF
#cross-references: GB:AE00743; NID:q2983675; PDB:q2983684; GB:AF006657
#experimental_source: strain VFS

GENETICS
#gene: aq1501
#summary: #length: 256 #molecular-weight: 29513 #checksum: 2157

Query Match 77.2% Score 44; DB 2; Length 256;
Best Local Similarity 44.4% Pred. No. 3.76e+01;
Matches 4; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

Db 28 SVKRISEVV 36
QY 1 SKRISVNV 9

```

```

**residues      1-457 **label KLE
**cross-references GB:AE 125; GB:AE000783; NID:g2688021;
                  PIDN:AE.56531.1; PID:g2688030; TIGR:BB0135
**experimental_source strain B31
CLASSIFICATION  *superfamily histidine--trna ligase; amino acid--trna ligase
                  *repeat homology; histidine--trna ligase; protein biosynthesis
KEYWORDS        aminoacyl--trna synthetase; ligase; protein biosynthesis
FEATURE         8-418
SUMMARY         *domain histidine--trna ligase homology *label H1L
                  *length 457 *molecular-weight 52848 *checksum 4898

Query Match      77.2% Score 44; DB 1; Length 457;
Best Local Similarity 56.7% Pred. No. 3,76e-01;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 246 SKRVEDVF 254
QY 1 SKRVSNV 9

RESULT 15
ENTRY   *type complete
TITLE   Probable NAD- ADP-ribosyltransferase (EC 2.4.2.30) - maize
ORGANISM *normal_name 24. - yys *common_name maize
DATE     24-Mar-1999 *sequence_revision 24-Mar-1999 *text_change
24-Mar-1999

ACCESSIONS  T03656
REFERENCE    Z14991
AUTHORS      Babyichuk, E.; Cottrill, P.; Storozhenko, S.; Fuanthiona, M.;
              O'Farrell, M.; Van Montagu, M.; Inze, D.; Kushnir, S.
SUBMISSION  Submitted to the FBL Data Library, November 1997
DESCRIPTION Higher plants possess two poly(ADP-ribose) polymerases.
ACCESSION   T03656
STATUS      preliminary; translated from GB/EMBL/CDR3J
**molecule_type mRNA
**residues 1-653 **label BAB
**cross-references EMBL:AJ222588; NID:e1264090; PID:e1264091;
GENETICS
GENE       PARP
KEYWORDS   DNA binding; glycosyltransferase; NAD; pentosyltransferase
SUMMARY    *length 653 *molecular-weight 72995 *checksum 7074

Query Match      77.2% Score 44; DB 2; Length 653;
Best Local Similarity 62.5% Pred. No. 3,76e-01;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 352 LKRSDVI 359
QY 2 IKRVSNV 9

Search Completed: Mon Jun 19 16:18:53 2000
Job time : 6 secs.

```

---

```

S66256      *type complete
alpha-1,6-mannosyl-glycoprotein beta-1,
2-N-acetylglucosaminyltransferase (EC 2.4.1.143) - human
*normal_name Homo sapiens *common_name man
28-Oct-1995 *sequence_revision 13-Mar-1997 *text_change
07-May-1999
ACCESSIONS  S66256
REFERENCE    S66256
AUTHORS      Tan, J.; d'Ayostaro, G.A.F.; Bendjak, B.; Reck, F.; Sarkar,
              M.; Squire, J.A.; Leong, P.; Schachter, H.
JOURNAL     Eur. J. Biochem. (1995) 231:317-328
TITLE       The human UDP-N-acetylglucosamine:alpha-6-D-mannoside-beta-1,
              2-N-acetylglucosaminyltransferase II gene (MGAT2). Cloning
              of genomic DNA, localization to chromosome 14q21,
              expression in insect cells and purification of the
              recombinant protein.
**cross-references M01D:95351854
ACCESSION   S66256
**molecule_type DNA
**residues 1-447 **label TAN
**cross-references EMBL:U15128; NID:g902744; PID:g902745
GENETICS
GENE       GDB:MGAT2
**cross-references GDB:137087
MAP_POSITION 14q21-14q21
KEYWORDS   glycoprotein; glycosyltransferase; Golgi apparatus;
              hexosyltransferase; transmembrane protein
FEATURE    10-28
59-86      *domain transmembrane *status predicted *label TMN
*binding-site carbohydrate (Asn) (covalent) *status
              predicted
SUMMARY    *length 447 *molecular-weight 51550 *checksum 8407

Query Match      77.2% Score 44; DB 2; Length 447;
Best Local Similarity 75.3% Pred. No. 3,76e-01;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 63 GIKRVSNV 70
QY 1 SKRVSNV 8

RESULT 14
ENTRY   *type complete
TITLE   Histidine--trna ligase (EC 6.1.1.21) hisS - Lyme disease
              Spirochaete
ALTERNATE_NAMES histidy--trna synthetase
ORGANISM *normal_name Borrelia burgdorferi *common_name Lyme disease
              Spirochaete
DATE     10-Sep-1999 *sequence_revision 10-Sep-1999 *text_change
10-Sep-1999
ACCESSIONS  G70116
REFERENCE    G70116
AUTHORS      Fraser, C.M.; Castens, S.; Huang, W.M.; Sutton, G.G.;
              Clayton, R.; Lathigra, R.; White, O.; Ketchum, K.A.;
              Dodson, R.; Hickey, E.K.; Gwinn, M.; Dougherty, B.; Tomb,
              J.F.; Fleischmann, R.D.; Richardson, D.; Peterson, J.;
              Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.;
              Vugt, R.V.; Palmer, N.; Adams, M.D.; Gocayne, J.; Weidman,
              J.; Uterback, L.; Wathey, L.; McDonald, L.; Artiach, P.;
              Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst,
              K.; Roberts, K.; Hatch, B.; Smith, H.O.; Venter, J.C.
JOURNAL     Nature (1997) 390:580-586
TITLE       Genomic sequence of a Lyme disease spirochaete, Borrelia
              burgdorferi.
**cross-references M01D:98065943
ACCESSION   G70116
STATUS      preliminary; nucleic acid sequence not shown;
              translation not shown
**molecule_type DNA

```

\*\*\*\*\*  
 W E B  
 \*\*\*\*\*  
 (IM)

Release 3.1A John F. Collins, BioComputing Research Unit,  
 Copyright (c) 1993-1998 University of Edinburgh, U.K.  
 Distribution rights by Oxford Molecular Ltd

MFSrch\_Lp protein - protein database search using Smith-Waterman algorithm

Run on: Mon Jun 19 16:17:57 2000: MasPar time 3.44 Seconds  
 79,741 Million cell updates/sec

Tabular output not generated.

Title: >US-09-142-524A-14  
 Description: (1-9) from US09142524A.pep  
 Perfect Score: 57  
 Sequence: 1 S:KRVSNVI:9  
 Scoring table: PAM 150  
 Gap 15

Searched: 83857 seqs, 30454973 residues

Post-processing: Minimum Match 0%  
 Missing first 45 summaries

Database: swissprot38  
 SwissProt

Statistics: Mean 22.889; Variance 21.707; scale 1.053

Pred. No. is the number of results predicted by chance to have a  
 score greater than or equal to the score of the result being printed,  
 and is derived by analysis of the total score distribution.

SUMMARY:ES

| Result No. | Score | Query Match | Length | BB | ID         | Description             | Pred. No. |
|------------|-------|-------------|--------|----|------------|-------------------------|-----------|
| 1          | 53    | 94.0        | 324    | 1  | SRP_CRYJA  | SUGI BASIC PROTEIN PRE  | 4.46e-02  |
| 2          | 50    | 87.7        | 326    | 1  | PELENT     | PERIATE LIPASE PRECURSO | 3.17e-01  |
| 3          | 47    | 82.5        | 358    | 1  | Y259_MYCPN | HYPOTHETICAL PROTEIN M  | 1.57e-00  |
| 4          | 44    | 77.2        | 142    | 1  | NK_AQUAE   | NUCLEOSIDE DIPHOSPHATE  | 1.02e-01  |
| 5          | 44    | 77.2        | 169    | 1  | RSS_CVAPA  | CYANELLE 7CS RIBOSOMAL  | 1.02e-01  |
| 6          | 44    | 77.2        | 173    | 1  | RSS_SYNY3  | 30S RIBOSOMAL PROTEIN   | 1.02e-01  |
| 7          | 44    | 77.2        | 195    | 1  | FUMB_METCA | PUTATIVE FUMARATE HYDR  | 1.02e-01  |
| 8          | 44    | 77.2        | 447    | 1  | GNT2_HUMAN | ALPHA-1,6-MANNOSYL-GLY  | 1.02e-01  |
| 9          | 44    | 77.2        | 457    | 1  | SHR_BORBU  | HISTIDYL-TRNA SYNTHETA  | 1.02e-01  |
| 10         | 44    | 77.2        | 611    | 1  | MEN1_MOUSE | MENIN.                  | 1.02e-01  |
| 11         | 44    | 77.2        | 615    | 1  | MEN1_HUMAN | MENIN.                  | 1.02e-01  |
| 12         | 44    | 77.2        | 720    | 1  | YM52_YEAST | HYPOTHETICAL 82.1 KD P  | 1.02e-01  |
| 13         | 44    | 77.2        | 968    | 1  | Y582_METCA | HYPOTHETICAL PROTEIN M  | 1.02e-01  |
| 14         | 43    | 75.4        | 102    | 1  | H4_PHYPO   | HISTONE H4              | 1.75e-01  |
| 15         | 43    | 75.4        | 174    | 1  | RSS_PORPU  | CHLOROPLAST 30S RIBOSO  | 1.75e-01  |
| 16         | 43    | 75.4        | 217    | 1  | RSS_MYCLE  | 30S RIBOSOMAL PROTEIN   | 1.75e-01  |
| 17         | 43    | 75.4        | 220    | 1  | RSS_MYCTU  | 30S RIBOSOMAL PROTEIN   | 1.75e-01  |
| 18         | 43    | 75.4        | 318    | 1  | KPR2_YEAST | RIBOSE-PHOSPHATE PRP    | 1.75e-01  |
| 19         | 43    | 75.4        | 350    | 1  | APBE_SALTY | THIAMINE BIOSYNTHESIS   | 1.75e-01  |
| 20         | 43    | 75.4        | 377    | 1  | Y2JH_BACSU | HYPOTHETICAL 42.0 KD P  | 1.75e-01  |
| 21         | 43    | 75.4        | 431    | 1  | AK_CHLIR   | ASPARTOKINASE (EC 2.7.  | 1.75e-01  |
| 22         | 43    | 75.4        | 730    | 1  | KOGA_MOUSE | DIACYLGLYCEROL KINASE,  | 1.75e-01  |
| 23         | 43    | 75.4        | 784    | 1  | LON_RICPR  | ATP-DEPENDENT PROTEASE  | 1.75e-01  |

RESULT: 1  
 ID SBP\_CRYJA STANDARD: PRI: 374 AA.  
 AC P18632;  
 DT 01-NOV-1990 (Rel. 16, Createc  
 DT 01-NOV-1995 (Rel. 32, Last se  
 DT 15-DEC-1998 (Rel. 37, Last an  
 DE SUGI BASIC PROTEIN PRECURSOR (SBP) (MAJOR ALLERGEN CRY 2.1) (CRY 2.1)  
 OS Cryptomeria japonica (Japanese cedar);  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC euphyllophytes; Spermatophyta; Coniferopsida; Coniferales;  
 OC Taxodiaceae; Cryptomeria.  
 RN [1]  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RC TISSUE+POLLEN;  
 RX MEDLINE: 94183234;  
 RA Sone T., Komiya N., Shimizu K., Kusakabe T., Morikubo K.,  
 RA Kino K.;  
 RT "Cloning and sequencing of cDNA coding for Cry 2.1, a major allergen  
 of Japanese cedar pollen";  
 RL Biochem. Biophys. Res. Commun. 199:619-625(1994).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE+POLLEN;  
 RA Nishida M., Kuroki Y., Iizuka K., Fukuda S., Kikimoto M.,  
 RN Submitted (1994); in the EMBL/GenBank/DDBJ databases  
 RP SEQUENCE OF 22-41;  
 RC TISSUE+POLLEN;  
 RX MEDLINE: 89031257;  
 RA Tanai M., Ando S., Usui M., Kiritomo M., Sakauchi M., Inouye S.,  
 RA Matsubashi T.;  
 RT "N-terminal amino acid sequence of a major allergen of Japanese cedar  
 pollen (Cry 2.1)";  
 RL FEBS Lett. 239:329-332(1988).  
 RN [4]  
 RP CARBOHYDRATES  
 RX MEDLINE: 95003748;  
 RA Hijikata A., Matsumoto I., Kojima K., Ogawa H.;  
 RT "Antigenicity of the oligosaccharide moiety of the Japanese cedar  
 (Cryptomeria japonica) pollen allergen, Cry 2.1";  
 RL Int. Arch. Allergy Immunol. 105:198-202(1994).  
 CC -!- PTM: CONTAINS FUCOSE/XULOSE-CONTAINING N-LINKED OLIGOSACCHARIDES.  
 CC -!- DISEASE: THIS PROTEIN IS THE MAJOR ALLERGEN OF JAPANESE CEDAR  
 CC -!- POLLEN: THE MOST COMMON POLLEN ALLERGEN IN JAPAN.  
 CC -!- MISCELLANEOUS: THE SEQUENCE OF CRY 2.1 FORM A IS SHOWN HERE. FORM  
 B DIFFERS IN SIX POSITIONS.  
 CC -!- SIMILARITY: BELONGS TO THE POLYSACCHARIDE LYASE FAMILY 1.

AMB A.1/AMB A.11/CRY J.1 SUBFAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL: D26544; AAA05542.1; 1  
 EMBL: D26545; AAA05543.1; 1  
 EMBL: D34699; AAA07020.1; 1  
 PIR: A44773; A44773  
 PFAM: PF00544; pec\_lyase; 1  
 PRINTS: PR00807; AMBALERGEN  
 Allergen: Glycoprotein; Multigene family; Signal;  
 FT SIGNAL 1 21  
 FT CHAIN 22 374 SUGI BASIC PROTEIN.  
 FT VARIANTE 12 12 L -> F (IN CRY J.1-B).  
 FT VARIANTE 143 143 H -> Y (IN CRY J.1-B).  
 FT VARIANTE 202 202 S -> T (IN CRY J.1-B).  
 FT VARIANTE 221 222 L -> S (IN CRY J.1-B).  
 FT VARIANTE 338 338 Q -> H (IN CRY J.1-B).  
 FT VARIANTE 354 361 K -> Q (IN CRY J.1-B).  
 FT VARIANTE 188 158 POTENTIAL.  
 FT CARBOHYD 191 191 POTENTIAL.  
 FT CARBOHYD 293 293 POTENTIAL.  
 SEQUENCE 174 AA: 40645 MW: 74825950248556 CRC64:  
 Query Match 93.08: Score 53: DB 1: Length 374:  
 Best Local Similarity 100.00: Pred. No. 4,96e-02:  
 Matches 8: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

Db 131 IKRVSNVI 138  
 QY 2 IKRVSNVI 9

RESULT 2 STANDARD: PRT: 426 AA.

ID PELEMENT  
 AC Q00845: 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE PECTATE LYASE PRECURSOR (EN 4.2.2.2).  
 GN PELA  
 OS Emericella nidulans (Aspergillus nidulans).  
 OC Eukaryota; Fungi; Ascomycota; Pleiomycetes; Emericellales;  
 IC Trichocommatales; Emericellaceae;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=FGSC 4;  
 RX MEDLINE: 95308556.  
 RA HO M.C., Whitehead M.P., Cleveland T.E., Dear R.A.;  
 RT "Sequence analysis of the Aspergillus nidulans pectate lyase peA  
 RI gene and evidence for binding of promoter regions to CREA, a  
 RL regulator of carbon catabolite repression.";  
 RL Curr. Genet. 27:142-149(1995).  
 CC 1- CATALYTIC ACTIVITY: ELIMINATIVE CLEAVAGE OF PECTATE TO GIVE  
 CC GLYCOSACCHARIDES WITH 4-DEOXY-ALPHA-D-MANN-4-ENURONSYL GROUPS  
 CC AT THEIR NON-REDUCING ENDS.  
 CC 1- SUBCELLULAR LOCATION: SECRETED (POTENTIAL).  
 CC 1- SIMILARITY: BELONGS TO THE POLYSACCHARIDE LYASE FAMILY 1.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC at the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL: U05592; AAA80568.1; 1  
 DR PFAM: PF00544; pec\_lyase; 1  
 KW Lyase; Signal; 1 15 POTENTIAL  
 FT SIGNAL 16 326 PECTATE LYASE.  
 FT CHAIN 16 326  
 SO SEQUENCE 326 AA: 34580 MW: 76445A4A5D615049 CRC54:  
 Query Match 87.78: Score 50: DB 1: Length 426:  
 Best Local Similarity 97.58: Pred. No. 3.17e-01:  
 Matches 7: Conservative 0: Mismatches 0: Indels 0:

Db 111 IKRVSNVI 113  
 QY 2 IKRVSNVI 9

RESULT 3 STANDARD: PRT: 438 AA.

ID Y269\_MYCPN  
 AC Y25395:  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE HYPOTHETICAL PROTEIN W3269 HXM.350.  
 OS Mycoplasma pneumoniae.  
 OC Bacteria; Firmicutes; Bacillus/clostridium group; Mollicutes.  
 CC Mycoplasmataceae; Mycoplasma.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 29342 / M129.  
 RX MEDLINE: 97165885  
 RA Himmelreich R., Hilbert H., Flanagan H., Farkl E., Li E.,  
 RA Herrmann R.;  
 RI "Complete sequence analysis of the genome of the bacterium Mycoplasma  
 RI pneumoniae.";  
 RL Nucleic Acids Res. 24:4420-4449(1996).  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC at the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL: AE000044; AAB96098.1; 1  
 KW Hypothetical protein.  
 SO SEQUENCE 348 AA: 42617 MW: A18E0FA248F15D03 CRC64:  
 Query Match 82.88: Score 41: DB 1: Length 426:  
 Best Local Similarity 77.88: Pred. No. 1.77e-02:  
 Matches 7: Conservative 0: Mismatches 0: Indels 0:

Db 349 SIKRVSKLI 357  
 QY 1 SIKRVSNVI 9

RESULT 4 STANDARD: PRT: 142 AA.

ID NDK\_AQUAE  
 AC G67528:  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 15-FEB-2000 (Rel. 39, Last annotation update)  
 DE NUCLEOSIDE DIPHOSPHATE KINASE (EC 2.7.4.6) (NDK) (NDP KINASE).  
 GN NDK OR AQ.1590.  
 OS Aquifex aeolicus.  
 OC Bacteria; Aquificales; Aquificaceae; Aquifex.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=VF5;  
 RX MEDLINE: 98196666.  
 RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.,  
 RA Graham D.E., Overbeek R., Sneed M.A., Keller M., Anjey M., Ruter P.,

RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.:  
 R1 "The complete genome of the hyperthermophilic bacterium Aquifex  
 R2 aeolicus".  
 R3 Nature 392:353-358(1998).  
 CC -1- FUNCTION: MAJOR ROLE IN THE SYNTHESIS OF NUCLEOSIDE TRIPHOSPHATES  
 CC OTHER THAN ATP.  
 CC -1- CATALYTIC ACTIVITY: ATP - NUCLEOSIDE DIPHOSPHATE - ADP -  
 CC NUCLEOSIDE TRIPHOSPHATE.  
 CC -1- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.  
 CC -1- SIMILARITY: BELONGS TO THE NDK FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: AEO0746; AAC07481.1; .  
 CC PROSITE: PS00462; NDP\_KINASES; 1.  
 CC PRAM: PF00334; NDK; 1.  
 CC Transferase: Kinase; ATP-binding.  
 CC ACT\_SITE 120 120 BY SIMILARITY.  
 CC SEQUENCE 142 AA: 18942 MW: 18985070.17E998 CRC64:  
 Query Match 77.2% Score 44: DR 1: Length 142:  
 Best Local Similarity 55.6% Pred. No. 1.02e-01:  
 Matches 5: Conservative 3: Mismatches 1: Indels 0: Gaps 0:  
 Db 22 AIRKRVREI 90  
 QY 1 SIKRVSNVI 9  
 RESULT 5  
 ID RS5\_CVAPA STANDARD: PRT: 169 AA.  
 AC P23402:  
 DT 01-NOV-1997 (Rel. 20, Created)  
 DI 01-NOV-1991 (Rel. 23, Last sequence update)  
 DE 01-FEB-1996 (Rel. 33, Last annotation update)  
 DE CYANELLE 3CS RIBOSOMAL PROTEIN S5.  
 GN RPS5.  
 OS Cyanophora paradoxa.  
 OC Eukaryota: Glaucocystophyceae: Cyanophoraceae: Cyanophora.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN: JTEX 5553.  
 RX MEDLINE: 9117154.  
 RA Michalowski C.B., Pflanzagl B., Loefelhardt W., Bohnert H.J.:  
 RT "The cyanelle S-10 spec ribosomal protein gene operon from Cyanophora  
 RC paradoxa".  
 RL Mol. Gen. Genet. 224:222-231(1990).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN: JB555 / PRINGSHEIM:  
 RA Stirewalt V.L., Michalowski C.B., Luffelhardt W., Bohnert H.J.:  
 RA Bryant D.A.:  
 RL Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: BELONGS TO THE SSP FAMILY OF RIBOSOMAL PROTEINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: M30487; AAA63629.1; .  
 CC EMBL: J30921; AAA81219.1; .

DR PIR: S12220; R3KTS.  
 DR HSSP: P02357; LPKP.  
 DR MENDEL: 7927; CYAPA: rps5.1.  
 DR PROSITE: PS00585; RIBOSOMAL\_S5; 1.  
 DR PRAM: PF00333; Ribosomal\_S5; 1.  
 KW Ribosomal protein; Cyanelle.  
 SQ SEQUENCE 169 AA: 17590 MW: 70533906.69812A9A CRC64:  
 Query Match 77.2% Score 44: DR 1: Length 169:  
 Best Local Similarity 62.5% Pred. No. 1.02e-01:  
 Matches 5: Conservative 4: Mismatches 0: Indels 0: Gaps 0:  
 Db 24 IRRVSKV 31  
 QY 2 IKRVSNVI 9  
 RESULT 6  
 ID RS5\_SYN3 STANDARD: PRT: 173 AA.  
 AC P71304:  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DI 01-NOV-1997 (Rel. 35, Last sequence update)  
 DE 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE 3CS RIBOSOMAL PROTEIN S5.  
 GN RPSE OF RPS5 OR SLL1812.  
 OS Synechocystis sp. (strain PC9803).  
 OC Bacteria: Cyanobacteria: Chroococcales: Synechocystis.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 97061201.  
 RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,  
 RA Miyajima N., Hiroseawa M., Sugita M., Sasamoto S., Kimura T.,  
 RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K.,  
 RA Okumura S., Shimpo S., Takeuchi C., Wada T., Watanabe A.,  
 RA Yamada M., Yasuda M., Tabata S.:  
 RT "Sequence analysis of the genome of the unicellular cyanobacterium  
 RT Synechocystis sp. strain PC9803. II. Sequence determination of the  
 RT entire genome and assignment of potential protein-coding regions".  
 RL DNA Res. 3:109-136(1996).  
 CC -1- FUNCTION: PROTEIN S5 IS IMPORTANT IN THE ASSEMBLY AND FUNCTION OF  
 CC THE 3CS RIBOSOMAL SUBUNIT (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE SSP FAMILY OF RIBOSOMAL PROTEINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: D90940; EAA17431.1; .  
 CC HSSP: P02357; LPKP.  
 DR PROSITE: PS00585; RIBOSOMAL\_S5; 1.  
 DR PRAM: PF00333; Ribosomal\_S5; 1.  
 KW Ribosomal protein.  
 SQ SEQUENCE 173 AA: 16241 MW: A930CF86E88475B1 CRC64:  
 Query Match 77.2% Score 44: DR 1: Length 173:  
 Best Local Similarity 62.5% Pred. No. 1.02e-01:  
 Matches 5: Conservative 3: Mismatches 0: Indels 0: Gaps 0:  
 Db 24 IRRVSKV 31  
 QY 2 IKRVSNVI 9  
 RESULT 7  
 ID F0MB\_METJA STANDARD: PRT: 195 AA.  
 AC Q58034:  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DI 01-NOV-1997 (Rel. 35, Last sequence update)  
 DE 01-NOV-1997 (Rel. 35, Last annotation update)

DE PUTATIVE FUMARATE HYDROLYTASE BETA SUBUNIT (EC 4.2.1.2) (FUMARASE).  
 GN MD0617.  
 OS Methanococcus jannaschii.  
 OC Archaea: Euryarchaeota: Methanococcales: Methanococcaceae:  
 CC Methanococcus.  
 RN [1].  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-JAL-1 / DSM 2661 / ATCC 43067;  
 RX MEDLINE: 93337999.  
 RA Bult C.J., White G., Olsen G.J., Zhou L., Fleischmann R.D.,  
 RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,  
 RA Kervatave A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.L.,  
 RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Gilek A.,  
 RA Scott C.L., Goodbagen N.S.M., Weidman C.F., Fushman D.L., Nguyen D.,  
 RA Uitterback L.E., Kelley J.M., Peterson S.D., Sadow P.W., Hanna M.C.,  
 RA Cotton M.D., Roberts K.M., Harst M.A., Kalne B.P., Borodovsky M.,  
 RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.:  
 RI "Complete genome sequence of the methanogenic archaeon, Methanococcus  
 RI jannaschii".  
 RL Science 273:1508-1573(1995).  
 CC -1- CATALYTIC ACTIVITY: L-MALATE + FUMARATE + H(2)O.  
 CC -1- PATHWAY: TRICARBOXYLIC ACID CYCLE.  
 CC -1- SIMILARITY: TO THE C-TERMINAL OF CLASS I FUMARASES.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL: G57510; AAF06121;  
 DP IGR: M0617;  
 KW Hypothetical protein; Hydase; Tricarboxylic acid cycle.  
 FT ACT\_SITE 35 36 POTENTIAL.  
 FT BINDING 104 104 CARBOXYL GROUP (POTENTIAL).  
 FT SEQUENCE 195 AA: 21866 MW: 4502607700502 CRC64:  
 SQ  
 Query Match 77.28; Score 44; DB 1; Length 195;  
 Best Local Similarity 75.04; Pred. No. 1.02e-01;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 Db 136 SVKRVNV 143  
 QY 1 SVKRVNV 8  
 RESULT 8  
 ID GN2\_EUKAN STANDARD: PRT: 447 AA.  
 AC Q10459;  
 DT 01-OCT-1995 (Rel. 34, Created)  
 DI 01-OCT-1996 (Rel. 34, Last sequence update)  
 DI 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE ALPHA-1,6-MANNOSYL-GLYCOPROTEIN BETA-1,2-N-  
 DE ACETYLGLUCOSAMINYLTRANSFERASE (EC 2.4.1.43) (N-GLYCOSYL-  
 DE GLUCOSACCHARIDE-GLYCOPROTEIN N-ACETYLGLUCOSAMINYLTRANSFERASE II)  
 DE (BETA-1,2-N-ACETYLGLUCOSAMINYLTRANSFERASE II) (GN1-II) (GLCNAC-T II).  
 GN Homo sapiens (Human).  
 OS Homo sapiens (Human).  
 OC Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Mammalia:  
 OC Eutheria: Primates: Catarrhini: Homnidae: Homo.  
 RN [1]  
 RP SEQUENCE FROM N.A. AND SEQUENCE OF 1-16.  
 RC TISSUE-LEUKOCYTE;  
 RX MEDLINE: 93361854.  
 RA Ian J., D'Agostaro A.F., Bendiak B., Reck F., Sarkar M., Squire J.A.,  
 RA Leong P., Schachter H.;  
 RI "The human UDP-N-acetylglucosamine: alpha-6-D-mannoside-beta-1,2-N-  
 RI acetylglucosaminyltransferase II gene (MGAT2). Cloning of genomic  
 RI DNA, localization to chromosome 14q21, expression in insect cells and  
 RI purification of the recombinant protein."  
 RL Eur. J. Biochem. 231:317-328(1995).

RN [2]  
 RP VARIANTS CG2 ARG-262 AND PHE-290.  
 RX MEDLINE: 96404413.  
 RA Tan J., Dunn J., Jaeken J., Schachter H.;  
 RI "Mutations in the MGAT2 gene controlling complex N-glycan synthesis  
 RI cause carbohydrate-deficient glycoprotein syndrome type II, an  
 RI autosomal recessive disease with defective brain development".  
 RL Am. J. Hum. Genet. 59:810-817(1996).  
 CC -1- FUNCTION: CATALYZES AN ESSENTIAL STEP IN THE CONVERSION OF GLC3  
 CC MANNOSE TO COMPLEX N-GLYCANS.  
 CC -1- CATALYTIC ACTIVITY: UDP-N-ACETYL-D-GLUCOSAMINE + ALPHA-D-MANN SYL-  
 CC 1,6-(N-ACETYL-BETA-D-GLUCOSAMINYL-1,2-ALPHA-D-MANNOSYL-1,4)-BETA-  
 CC D-MANNOSYL-1,6-(N-ACETYL-BETA-D-GLUCOSAMINYL-1,2-ALPHA-1,4)-BETA-  
 CC 1,3)-BETA-D-MANNOSYL-R  
 CC -1- PATHWAY: GLYCOSYLATION.  
 CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. GOLG.  
 CC -1- DISEASE: CONGENITAL DEFICIENCY OF MGAT2 IS ASSOCIATED WITH SEVERE  
 CC IMPAIRMENT OF NORMAL EMERYGNESEN, PARTICULARLY IN THE NERVOUS  
 CC SYSTEM IN PATIENTS WITH CARBOHYDRATE-DEFICIENT GLYCOPROTEIN  
 CC SYNDROME II (CDGS TYPE II OR CG2). IT HAS ALSO BEEN IMPLICATED IN  
 CC THE ETIOLOGY OF A VARIANT FORM OF CONGENITAL DYSENRYTHROGENETIC  
 CC ANEMIA TYPE II.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL: U15128; AAA86956.1;  
 DR MIM: 60265.6;  
 DR MIM: 212066;  
 KW Transferase: Glycosyltransferase; Transmembrane; Signal-anchor.  
 KW Glycoprotein; Golgi stack; Disease mutation.  
 FT DOMAIN 1 9 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 10 29 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).  
 FT DOMAIN 30 447 LUMENAL, CATALYTIC (POTENTIAL).  
 FT CARBOHYD 59 59 POTENTIAL.  
 FT CARBOHYD 86 86 POTENTIAL.  
 FT VARIANT 252 252 H->R (IN CG2).  
 FT VARIANT 290 290 /FTID-VAR\_003425.  
 FT VARIANT 290 290 S->F (IN CG2).  
 FT /FTID-VAR\_003416.  
 FT /FTID-VAR\_003416.  
 SQ SEQUENCE 447 AA: 51550 MW: 54307608808572 CRC64:  
 Query Match 77.28; Score 44; DB 1; Length 447;  
 Best Local Similarity 75.04; Pred. No. 1.02e-01;  
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
 Db 6 SVKRVNV 70  
 QY 1 SVKRVNV 8  
 RESULT 9  
 ID SYH\_BORBU STANDARD: PRT: 457 AA.  
 AC O51160;  
 DI 15-DEC-1998 (Rel. 37, Created)  
 DI 15-DEC-1998 (Rel. 37, Last sequence update)  
 DI 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE HISTIDYL-TRNA SYNTHETASE (EC 6.1.1.21) (HISTIDINE--TRNA LIGASE);  
 DE (HISRS).  
 GN HIS OR BBO135.  
 OS Borrelia burgdorferi (Lyme disease spirochete).  
 OC Bacteria: Spirochaetales: Spirochaetaceae: Borrelia.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-ATCC 35210 / B31;  
 RX MEDLINE: 98065943.  
 RA Fraser C.M., Casjens S., Huang W.M., Sutton G.G., Clayton R.A.,



RA Latnigra R., White C., Ketchum K.A., Dodson R., Hickey E.K., Gwinn M.,  
RA Dougherty B., Tomb J.-F., Fleischmann R.D., Richardson D.,  
RA Peterson J., Kerlavage A.R., Quackenbush J., Salzberg S., Hanson M.,  
RA Van Vugt R., Palmer N., Adams M.D., Gocayne J.D., Weidman J.,  
RA Utterback T., Watney L., McDonald L., Attilach P., Bowman C.,  
RA Garland S., Fujii C., Cotton M.D., Horst K., Roberts K., Hatch B.,  
RA Smith H.O., Venter J.C.,  
RT "Genomic sequence of a Lyme disease spirochete, Borrelia  
RT burgdorferi,"  
RL Nature 392:580-595(1992).  
CC -- CATALYTIC ACTIVITY: ATP -> L-HISTIDINE + TRNA(HIS) - AMP +  
CC -- PYROPHOSPHATE -> L-HISTIDYL-TRNA(HIS).  
CC -- SUBUNIT: HOMODIMER (BY SIMILARITY).  
CC -- SUBCELLULAR LOCATION: CYTOPLASMIC.  
CC -- SIMILARITY: BELONGS TO CLASS-II AMINOACYL-TRNA SYNTHETASE FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation  
CC at the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: AF001125; AAC6531.1;  
DR EMBL: AF001125; AAC6531.1;  
DR EMBL: AF016398; AAC79938.1;  
KW Nuclear protein.  
FT CONFLICT 457 457 M -> I (IN REF. 2).  
FT CONFLICT 466 466 G -> E (IN REF. 2).  
FT CONFLICT 512 512 I -> S (IN REF. 2).  
SQ SEQUENCE 611 AA; 67472 MW; C3FBB3A5655244E9 CRC64.  
  
Query Match 77.28; Score 44; DB 1; Length 457;  
Best Local Similarity 65.78; Pred. No. 1.02e+01;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
  
DB 246 SIKRVSDVF 254  
QY I I I I I I  
QY I SIKRVSNVI 9  
  
RESULT 10  
ID MEN1\_MOUSE STANDARD: PRI: 611 AA.  
AC O88559;  
DI 15-DEC-1998 (Rel. 37, Created);  
DI 15-DEC-1998 (Rel. 37, Last sequence update);  
DI 15-FEB-2000 (Rel. 39, Last annotation update);  
DE MEN1.  
GN Mus musculus (Mouse).  
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
RN [1]  
RC SEQUENCE FROM N.A. AND VARIANTS MEN1.  
RX MEDLINE: 97258940.  
RA Chandrasekharappa S.C., Guru D., Manickam P., Olufemi S.E.,  
RA Collins F.S., Emmert-Buck M., Debelenko L.V., Zhuang Z.,  
RA Lubensky I.A., Liotta L.A., Green J.S., Guru S.C., Manickam P.,  
RA Weisman J., Boquski M.S., Iwai S.K., Kester M.B., Kim Y.S.,  
RA Heppner C., Dong Q., Spiegel A.M., Burns A.L., Marx S.J.,  
RA "Positional cloning of the gene for multiple endocrine neoplasia-type  
RA 1,"  
RL Science 276:434-407(1997).  
RN [2]  
RN Science 276:434-407(1997).  
RP VARIANTS FMEN1 AND SPURADIC MEN1.  
RX MEDLINE: 97385593.  
RA Agarwal S.K., Kester M.B., Debelenko L.V., Heppner C.,  
RA Emmert-Buck M.R., Skarulis M.C., Doppman J.L., Kim Y.S.,  
RA Lubensky I.A., Zhuang Z., Green J.S., Guru S.C., Manickam P.,  
RA Olufemi S.E., Liotta L.A., Chandrasekharappa S.C., Collins F.S.,  
RA Spiegel A.M., Burns A.L., Marx S.J.,  
RA "Germline mutations of the MEN1 gene in familial multiple endocrine  
RA neoplasia type 1 and related states,"  
RT Hum. Mol. Genet. 6:1169-1175(1997).  
RL [3]  
RN [3]  
RP VARIANT PARATHYROID ADENOMA LYS-26.  
RX MEDLINE: 97385243.  
RA Heppner C., Kester M.B., Agarwal S.K., Debelenko L.V.,  
RA Emmert-Buck M.R., Guru S.C., Manickam P., Olufemi S.E.,  
RA Skarulis M.C., Doppman J.L., Alexander R.H., Kim Y.S., Sagar S.K.,  
RA Lubensky I.A., Zhuang Z., Liotta L.A., Chandrasekharappa S.C.,  
RA Collins F.S., Spiegel A.M., Burns A.L., Marx S.J.,  
RA "Somatic mutation of the MEN1 gene in parathyroid tumours,"  
RT Nat. Genet. 16:375-378(1997).  
RL [4]  
RN [4]  
RP VARIANTS FMEN1 D-42; P-165; D-169; S-188 AND E-289.  
RX MEDLINE: 98130524.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation  
CC at the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: AF072755; AAC26001.1;  
DR EMBL: AF072755; AAC26001.1;  
DR EMBL: AF016398; AAC79938.1;  
KW Nuclear protein.  
FT CONFLICT 457 457 M -> I (IN REF. 2).  
FT CONFLICT 466 466 G -> E (IN REF. 2).  
FT CONFLICT 512 512 I -> S (IN REF. 2).  
SQ SEQUENCE 611 AA; 67472 MW; C3FBB3A5655244E9 CRC64.  
  
Query Match 77.28; Score 44; DB 1; Length 611;  
Best Local Similarity 62.58; Pred. No. 1.02e+01;  
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;  
  
DB 118 VKKVSQVI 125  
QY I I I I I I  
QY I I I I I I  
QY 2 I I I I I I  
  
RESULT 11  
ID MEN1\_HUMAN STANDARD: PRI: 615 AA.  
AC O00255; O00632;  
DI 15-DEC-1998 (Rel. 37, Created);  
DI 15-DEC-1998 (Rel. 37, Last sequence update);  
DI 15-FEB-2000 (Rel. 39, Last annotation update);  
DE MEN1.  
GN HOMO SAPIENS (Human).  
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
RN [1]  
RC SEQUENCE FROM N.A. AND VARIANTS FMEN1.  
RX MEDLINE: 97258940.  
RA Chandrasekharappa S.C., Guru D., Manickam P., Olufemi S.E.,  
RA Collins F.S., Emmert-Buck M., Debelenko L.V., Zhuang Z.,  
RA Lubensky I.A., Liotta L.A., Green J.S., Guru S.C., Manickam P.,  
RA Weisman J., Boquski M.S., Iwai S.K., Kester M.B., Kim Y.S.,  
RA Heppner C., Dong Q., Spiegel A.M., Burns A.L., Marx S.J.,  
RA "Positional cloning of the gene for multiple endocrine neoplasia-type  
RA 1,"  
RL Science 276:434-407(1997).  
RN [2]  
RN Science 276:434-407(1997).  
RP VARIANTS FMEN1 AND SPURADIC MEN1.  
RX MEDLINE: 97385593.  
RA Agarwal S.K., Kester M.B., Debelenko L.V., Heppner C.,  
RA Emmert-Buck M.R., Skarulis M.C., Doppman J.L., Kim Y.S.,  
RA Lubensky I.A., Zhuang Z., Green J.S., Guru S.C., Manickam P.,  
RA Olufemi S.E., Liotta L.A., Chandrasekharappa S.C., Collins F.S.,  
RA Spiegel A.M., Burns A.L., Marx S.J.,  
RA "Germline mutations of the MEN1 gene in familial multiple endocrine  
RA neoplasia type 1 and related states,"  
RT Hum. Mol. Genet. 6:1169-1175(1997).  
RL [3]  
RN [3]  
RP VARIANT PARATHYROID ADENOMA LYS-26.  
RX MEDLINE: 97385243.  
RA Heppner C., Kester M.B., Agarwal S.K., Debelenko L.V.,  
RA Emmert-Buck M.R., Guru S.C., Manickam P., Olufemi S.E.,  
RA Skarulis M.C., Doppman J.L., Alexander R.H., Kim Y.S., Sagar S.K.,  
RA Lubensky I.A., Zhuang Z., Liotta L.A., Chandrasekharappa S.C.,  
RA Collins F.S., Spiegel A.M., Burns A.L., Marx S.J.,  
RA "Somatic mutation of the MEN1 gene in parathyroid tumours,"  
RT Nat. Genet. 16:375-378(1997).  
RL [4]  
RN [4]  
RP VARIANTS FMEN1 D-42; P-165; D-169; S-188 AND E-289.  
RX MEDLINE: 98130524.

BA Bassett J.H.D., Forbes S.A., Pannett A.A.J., Lloyd S.E.,  
 RA Christie P.T., Wooding C., Harding B., Besser G.M., Edwards C.R.,  
 RA Monson J.P., Sampson J., Wass J.A.H., Wheeler M.H., Thakker R.V.,  
 RA "Characterization of mutations in patients with multiple endocrine  
 RA neoplasia type 1.",  
 RL Am. J. Hum. Genet. 62:232-244(1998).  
 RP [5]  
 RP VARIANTS FMEN1.  
 RP MEDLINE: 98149469.  
 RA Giraud S., Zhang C.X., Serova-Sinilnikova O., Wautot V., Salandre J.,  
 RA Buissone N., Waterloo C., Bauliers C., Porchet N., Aubert J.P., Ezy P.,  
 RA Cadot G., Briener B., Chabre C., Niccoli P., Lefrat F., Duron P.,  
 RA Empedador B., Goudet P., Sarfati E., Riou C.P.,  
 RA Guichard S., Rouder M., Meyrier A., Caron P., Vantghem M.C.,  
 RA Assayag M., Peix J.L., Puget M., Rohner V., Valotton M., Lenoir G.,  
 RA Gaudray P., Poye C., Conte-Devolx B., Chanson P., Stugart Y.Y.,  
 RA Goldgar D., Murat A., Calender A.,  
 RA "Germline mutation analysis in patients with multiple endocrine  
 RA neoplasia type 1 and related disorders.",  
 RL Am. J. Hum. Genet. 63:455-467(1998).  
 RP [6]  
 RP VARIANTS FMEN1.  
 RP MEDLINE: 99011276.  
 RA Teh B.T., Esapa C.T., Houston R., Grandell U., Parnebo F.,  
 RA Nordenskjold M., Pearce C.J., Carmichael D., Larsson C., Harris P.E.,  
 RA "A family with isolated hyperparathyroidism segregating a missense  
 RA MEN1 mutation and showing loss of the wild-type alleles in the  
 RA parathyroid tumors.",  
 RL Am. J. Hum. Genet. 63:1544-1549(1998).  
 RP [7]  
 RP VARIANTS FMEN1.  
 RP MEDLINE: 99077176.  
 RA Fujimori M., Shirahama S., Sakurai A., Hashizume K., Hana Y., Ito K.,  
 RA Shingu K., Kobayashi S., Amano J., Fukushima Y.,  
 RA "Novel V184E MEN1 germline mutation in a Japanese kindred with  
 RA familial hyperparathyroidism.",  
 RL Am. J. Med. Genet. 80:221-222(1996).  
 RP [8]  
 RP VARIANTS FMEN1.  
 RP MEDLINE: 98149442.  
 RA Agarwal S.K., Gabeleuko L.V., Kester M.B., Gura S.C., Manickam P.,  
 RA Olufemi S.E., Skarulis M.C., Heppner C., Trabtree J.S.,  
 RA Lubensky I.A., Zhuang Z., Kim Y.S., Chandrasekharappa S.C.,  
 RA Collins F.S., Liotta L.A., Spiegel A.M., Burns A., Emmert-Buck M.R.,  
 RA Marx S.J.,  
 RA "Analysis of recurrent germline mutations in the MEN1 gene encountered  
 RA in apparently unrelated families.",  
 RL Hum. Mutat. 12:75-82(1998).  
 RP [9]  
 RP VARIANTS FMEN1: 112-135 AND CYS 364  
 RP MEDLINE: 98430971.  
 RA Boeri R., Volpinetti A., Buck S., Park W.S., Kim G., Hofbauer B.,  
 RA Darlino J., Linde H., Zhuang Z.,  
 RA "Somatic mutations of the MEN1 tumor suppressor gene detected in  
 RA sporadic angeliobromas.",  
 RL J. Invest. Dermatol. 111:539-540(1998).  
 RP [10]  
 RP VARIANTS FMEN1: CYS-119 DEL AND GLN-171-LEU-173 DEL.  
 RP MEDLINE: 98439173.  
 RA Sakurai A., Shirahama S., Fujimori M., Katal M., Itakura Y.,  
 RA Kobayashi S., Amano J., Fukushima Y., Hashizume K.,  
 RA "Novel MEN1 gene mutations in familial multiple endocrine neoplasia  
 RA type 1.",  
 RL J. Hum. Genet. 43:199-201(1998).  
 RP [11]  
 RP VARIANT FMEN1 GLY-45.  
 RP MEDLINE: 99048878.  
 RA Sato M., Matsubara S., Miyauchi A., Ohye H., Imachi H., Murao K.,  
 RA Takahara J.,  
 RA "Identification of five novel germline mutations of the MEN1 gene in  
 RA Japanese multiple endocrine neoplasia type 1 (MEN1) families.",  
 RL J. Med. Genet. 35:915-919(1998).  
 RP [12]  
 RP VARIANTS FMEN1 TRP-39; TYR-177; ASP-184; PRO-269 AND PRO-272.  
 RP MEDLINE: 99103464.  
 RA Poncin J., Abs R., Velkeniers E., Bonduelle M., Abramowitz M.,  
 RA Legros J.-J., Verloes A., Meunasse M., van Gaal L., Verellen D.,  
 RA Koutischer L., Beckers A.,  
 RA "Mutation analysis of the MEN1 gene in Belgian patients with multiple  
 RA endocrine neoplasia type 1 and related diseases.",  
 RL Hum. Mutat. 13:54-60(1999).  
 RP [13]  
 RP VARIANTS MEN1 ASP-161 AND ARG 146.  
 RP MEDLINE: 99168381.  
 RA Katch M.G., Dilley W.G., Sanjurjo F., Deberedetti M.K., Roberts J.M.,  
 RA Wells S.A., Jr., Goodfellow P., Calimire T.C.,  
 RA "Germline mutations in the multiple endocrine neoplasia type 1 gene  
 RA evidence for frequent splicing defects.",  
 RL Hum. Mutat. 13:175-185(1999).  
 CC -- FUNCTION: NOT KNOWN.  
 CC -- SUBCELLULAR LOCATION: NUC.  
 CC -- TISSUE SPECIFICITY: URICAC.  
 CC -- DISEASE DEFECTS IN MEN1: THE CAUSE OF FAMILIAL MULTIPLE  
 CC ENDOCRINE NEOPLASIA TYPE 1 (MEN1); WERNER SYNDROME; AN AUTOSOMAL  
 CC DOMINANT DISORDER CHARACTERIZED BY TUMORS OF THE PARATHYROID  
 CC GLANDS, GASTRO-INTESTINAL ENDOCRINE TISSUE, THE ANTERIOR PITUITARY  
 CC AND OTHER TISSUES. CUTANEOUS LESIONS AND NERVOUS-TISSUE TUMORS  
 CC CAN EXIST. PROGNOSIS IN MEN1 PATIENTS IS RELATED TO HORMONAL  
 CC HYPERSECRETION BY TUMORS, SUCH AS HYPERGASTRINEMIA CAUSING SEVERE  
 CC PEPTIC ULCER DISEASE (ZOLLINGER-ELISON SYNDROME, ZES), PRIMARY  
 CC HYPERPARATHYROIDISM, AND ACUTE FORMS OF HYPERINSULINEMIA.  
 CC -- DISEASE DEFECTS IN MEN1: IS THE CAUSE OF FAMILIAL ISOLATED  
 CC HYPERPARATHYROIDISM (FHP OR HRPT). FHP IS AN AUTOSOMAL DOMINANT  
 CC DISORDER CHARACTERIZED BY HYPERCALCEMIA, ELEVATED PARATHYROID  
 CC HORMONE (PTH) LEVELS, AND UNIGLANDULAR OR MULTIGLANDULAR  
 CC PARATHYROID TUMORS.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL database, the  
 CC European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC EMBL: U91236; AAC51229.1;  
 CC EMBL: U91237; AAC51230.1;  
 CC MIM: 131000;  
 CC MIM: 145000;  
 KW Nuclear protein; Disease data;  
 FT VARSPLIC 149 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000

```

FT. VARIANT 161 161 G -> D (IN MEN1).
FT FTID=VAR_008017.
FT VARIANT 165 165 A -> P (IN FMEN1).
FT FTID=VAR_005437.
FT VARIANT 169 169 A -> D (IN FMEN1).
FT FTID=VAR_005438.
FT VARIANT 171 171 MISSING (IN FMEN1).
FT FTID=VAR_005439.
...
Note: remainder of annotations omitted.

Query Match 77.2% Score 44 DB 1: Length 615
Best Local Similarity 62.5% Pred. No. 1.02e-01
Matches 5: Conservative 3: Mismatches 0: Indels 0: Gaps 0:

DE 118 VKRSDV 125
QY 2 IKRVSNI 9

RESULT 12
ID YMS2-YEAST STANDARD: PRT: 720 AA.
AC Q04322.
DT 01-NOV-1997 (Rel. 35, Created).
DI 01-NOV-1997 (Rel. 35, Last sequence update).
DI 01-NOV-1997 (Rel. 35, Last annotation update).
DE HYPOTHETICAL 82.1 KD PROTEIN IN SGS-MRPL24 INTERGENIC REGION.
GN YMS192W OR YMS646.04.
OS Saccharomyces cerevisiae (Paker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetaceae;
OC Saccharomycetaceae; Saccharomycetes; Saccharomycetes;
RN 1.
RE SEQUENCE FROM N.A.
RC STRAIN=5288C / AB972.
RL Pearson D., Swman S., Harrell B.G., Rajandream M.A.;
RL SUBMITTED (JAN-1995) to the EMBL/GenBank/DBE databases.
CC -1- SIMILARITY: TO YEAST YPL249C AND S.POMBE SPAC26F1.09.
...
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@sib-sib.ch).
CC EMBL: 247815; CAA87131.1.
CC HYPOTHETICAL protein
KW Hypothetical protein
SQ SEQUENCE 720 AA: 82.1% MW: 81230849.57247 19.14

Query Match 77.2% Score 44 DB 1: Length 720
Best Local Similarity 62.5% Pred. No. 1.02e-01
Matches 5: Conservative 3: Mismatches 0: Indels 0: Gaps 0:

DE 352 VKRISNL 359
QY 2 IKRVSNI 9

RESULT 13
ID Y682-REITA STANDARD: PRT: 968 AA.
AC Q58095.
DT 01-NOV-1997 (Rel. 35, Created).
DI 01-NOV-1997 (Rel. 35, Last sequence update).
DI 15-DEC-1998 (Rel. 37, Last annotation update).
DE HYPOTHETICAL PROTEIN M30682.
GN M30682.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
OC Methanococcus.
RN 1.
RE SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;

```

```

RX MEDLINE: 96337999.
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich J.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Staden A.,
RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fushman J.D., Nguyen P.,
RA Cotton M.D., Roberts K.M., Hirst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.P., Fraser C.M., Smith H.G., Woese C.R., Venter J.C.
RI *Complete genome sequence of the methanogenic archaeon, Methanobrevibacter
RI jannaschii.
RL Science 273:1058-1073(1996)
CC -1- P1M: THIS PROTEIN UNDERGOES A PROTEIN SELF SPLICING THAT INVOLVES
CC A POST-TRANSLATIONAL EXCISION OF THE THE INTERVENING REGION
CC (INTER) FOLLOWED BY PEPTIDE LIGATION (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE UPF0027 (RTCB) FAMILY.
...
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@sib-sib.ch).
CC EMBL: J67515; AAB98677.1.
CC -1- GR: M30582.
DR PROSITE: PS00881; PROTEIN_SPLICING.
DR PROSITE: PS01288; UPF0027.1.
DR PFAM: PF01139; UPF0027.1.
KW Hypothetical protein; Protein splicing
FT CHAIN 1 97 HYPOTHETICAL PROTEIN M30682.1ST PART
FT (POTENTIAL).
FT CHAIN 98 585 M3A HYP2 INTER (POTENTIAL).
FT CHAIN 586 968 HYPOTHETICAL PROTEIN M30682.2ND PART
FT (POTENTIAL).
SQ SEQUENCE 968 AA: 110256 MW: 3982466260456130 CRC64.

Query Match 77.2% Score 44 DB 1: Length 968
Best Local Similarity 62.5% Pred. No. 1.02e-01
Matches 5: Conservative 3: Mismatches 0: Indels 0: Gaps 0:

DE 5 LKRVSDV 12
QY 2 IKRVSNI 9

RESULT 14
ID R41HVPV STANDARD: PRT: 100 AA.
AC P04915.
DI 13-AUG-1987 (Rel. 35, Created).
DI 13-AUG-1987 (Rel. 35, Last sequence update).
DI 15-JUL-1999 (Rel. 38, Last annotation update).
DE HISTONE H4.
GN H41 AND H42.
OS Physarum polycephalum (Slime mold).
OC Eukaryota; Myxogastria; Physarida; Physarum.
RN 1.
RE SEQUENCE FROM N.A.
RX MEDLINE: 85279930.
RA Wilhelm M.L., Wilhelm F.-X.
RI *Histone genes in Physarum polycephalum: transcription and analysis
RI of the flanking regions of the two H4 genes.
RL J. Mol. Evol. 28:322-326(1989).
RN 2.
RE SEQUENCE FROM N.A.
RX MEDLINE: 87260013.
RA Wilhelm M.L., Wilhelm F.-X.
RI *Both histone H4 genes of Physarum polycephalum are interrupted by an
RI intervening sequence.
RL Nucleic Acids Res. 15:5478-5478(1987).
RN 3.
RE SEQUENCE FROM N.A.

```

```

RX MEDLINE: 84208772.
RA Wilhelm M., Wilhelm F.X.:
RT "A transposon-like DNA fragment interrupts a Physarum polycephalum
RT histone H4 gene".
RL FEBS Lett. 458:249-254(1984).
CC -- FUNCTION: HISTONE H4 ALONG WITH HISTONE H3, PLAYS A CENTRAL ROLE
CC IN NUCLEOSOME FORMATION.
CC -- SUBUNIT: THE NUCLEOSOME IS AN OCTAMER CONTAINING TWO MOLECULES OF
CC H2A, H2B, H3, AND H4, WHICH WRAP APPROXIMATIVELY 146 BP OF DNA.
CC -- MISCELLANEOUS: THE SEQUENCE OF H41 IS SHOWN.
CC -- SIMILARITY: BELONGS TO THE HISTONE H4 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement. (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch.)
CC
CC EMBL: X00449; CAA25140.1;
D6 EMBL: X15141; CAA33239.1;
D6 EMBL: X15142; CAA33240.1;
D6 EMBL: X3036; CAA68442.1;
D6 PIR: A27859; A27859.
D6 PIR: S10075; S10075.
D6 PIR: S10076; S10076.
D6 PRINTS: PR00623; HISTONE_H4_1.
D6 PRFAM: PRF00247; HISTONE_H4_1.
D6 PRFAM: PRF00225; histone.
KW Chromosomal protein; Nucleosome core; Nuclear protein; DNA-binding.
FT INIT_MET 0
FT DNA_BIND 15
FT VARIANT 48 48 N -> K (IN H42).
FT VARIANT 49 48 N -> K (IN H42).
SV SEQUENCE 102 AA; 11311 MW; 8223795E72841FC CRC64;

```

Query Match 75.4% Score 43 DB 11 Length 102

Best Local Similarity 55.6% Pred. NO. 1.75e-01

Matches 5: conservative 4: Mismatches 1: Indels 0: Gaps 0:

Db 42 GVRKRNNT 50

111111

1 SKRVSNVI 9

```

RESULT 15
ID RH5_PORPY STANDARD: PRI: 174 AA.
AC P51248:
CI 01-01-1994 (Ref. 34, Cited)
CI 01-01-1994 (Ref. 34, Last sequence update)
CI 01-01-1994 (Ref. 34, Last sequence update)
DE CHLOROPLAST 30S RIBOSOMAL PROTEIN S6.
GN RPS5.
OS Porphyra purpurea.
OG Chloroplast.
OC Eukaryota; Rhodophyta; Bangiophyceae; Bangiales; Bangiaceae; Porphyra.
RN 1.
RP SEQUENCE FROM N.A.
RC STRAIN=AVONPORT.
RA Reith M.E., Munnholland J.:
RI "Complete nucleotide sequence of the Porphyra purpurea chloroplast
RI genome".
RL Plant Mol. Biol. Rep. 13:333-335(1995)
CC -- SIMILARITY: BELONGS TO THE S5P FAMILY OF RIB-SOMAL PROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement. (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch.)
CC

```

```

DR EMBL: U38804; AAC08184.1;
DR HSP: P02357; 1PKP.
DR MENDEL: 10355; PORPU:RPS5;1.
DR PRSITE: PS00585; RIBOSOMAL_S5; 1.
DR PRAM: PF00333; Ribosomal_S5;1
KW Ribosomal protein; Chloroplast
SC SEQUENCE 174 AA; 18294 MW; 9052BDA332242460C CRC64;

```

Query Match 75.4% Score 43 DB 11 Length 102  
Best Local Similarity 50.0% Pred. NO. 1.75e-01

Matches 4: conservative 4: Mismatches 0: Indels 0: Gaps 0:

Db 24 VRRVTKVV 31

111111

2 IKRVSNVI 9

Search completed: Mon Jun 19 16:23:37 2000

Job time : 6 secs.



OS Escherichia coli.  
 CG Plasmid Incil Colib-p9.  
 CC Bacteria: Proteobacteria, gamma subdivision; Enterobacteriaceae;  
 CC Escherichia.  
 RN [1].  
 RP SEQUENCE FROM N.A.  
 RA SAMPEL G., MIZUBUCHI K.:  
 RI "Organization and diversification of plasmid genomes: complete  
 RI nucleotide sequence of the Colib-p9 genome."  
 RJ Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AB021078; BAA76111;  
 KW Plasmid.  
 SQ SEQUENCE 326 AA: 36227 MW: 92469414 CRC32:

Query Match 84.2% Score 48: DB 2: Length 326;  
 Best Local Similarity 75.0% Pred. NO. 1.87e-00;  
 Matches 6: Conservative 2: Mismatches 0: Indels 0: Gaps 0:

DB 26: IKRVNVI 268  
 |||||  
 QY 2 IKRVNVI 9

RESULT 3 PRELIMINARY: PRI: 411 AA.  
 ID Q46322  
 AC Q46322  
 DT 01-JUN-1998 (TRENBLrel. 06, Created)  
 DI 01-JUN-1998 (TRENBLrel. 06, Last sequence update)  
 DE 01-AUG-1998 (TRENBLrel. 07, Last annotation update)  
 DE PLASMID GR04970, COMPLETE SEQUENCE.  
 CS Gracillaria robusta.  
 OS Plasmid GR04970.  
 CC Eukaryota; Rhodophyta; Florideophyceae; Gracillariales; Gracillariaceae;  
 CC Gracillaria.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA GOFF L.J., KERN D.A.:  
 RI Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF034718; AAC04724.1;  
 KW Plasmid.  
 SQ SEQUENCE 411 AA: 48427 MW: 63365E93 CRC32:

Query Match 80.7% Score 46: DB 10: Length 411;  
 Best Local Similarity 62.5% Pred. NO. 5.90e-00;  
 Matches 5: Conservative 3: Mismatches 0: Indels 0: Gaps 0:

DB 10: IKRVNVI 17  
 |||||  
 QY 2 IKRVNVI 9

RESULT 4 PRELIMINARY: PRI: 465 AA.  
 ID Q23157  
 AC Q23157  
 DT 01-NOV-1996 (TRENBLrel. 01, Created)  
 DI 01-JAN-1998 (TRENBLrel. 05, Last sequence update)  
 DE 01-MAY-1999 (TRENBLrel. 10, Last annotation update)  
 DE W04D2.3 PROTEIN.  
 GN W04D2.3.  
 OS Caenorhabditis elegans.  
 CC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;  
 CC Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA LENNARD N.:  
 RI Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.  
 DR [2]  
 RN SEQUENCE FROM N.A.  
 RX MEDLINE: 94150718.  
 RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,  
 RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,  
 RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAYELLO A., FULTON L.,  
 RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,

RA JONES M., KERSHAM J., KIRSTEN J., LAISTER N., LATREILLE P.,  
 RA LIGHTING J., LLOYD C., MCILRAY A., MORTIMORE B., O'CALLAGHAN M.,  
 RA PARSONS J., PERCY C., RIFKEL B., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,  
 RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,  
 RA THIERSY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,  
 RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT C., WOULDMAN P.,  
 RT \*2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
 RT elegans.\*  
 RL Nature 368:32-38(1994).  
 DR EMBL: Z75552; CAA99940.1;  
 DR PEARL: PF00876; Icnexin; 1.  
 SQ SEQUENCE 465 AA: 54576 MW: 2FA84759 CRC32:

Query Match 80.7% Score 46: DB 5: Length 465;  
 Best Local Similarity 100.0% Pred. NO. 5.90e-00;  
 Matches 7: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

DB 456 KRVSNVI 462  
 |||||  
 QY 3 KRVSNVI 9

RESULT 5 PRELIMINARY: PRI: 78: AA.  
 ID Q92H39  
 AC Q92H39  
 DT 01-MAY-1999 (TRENBLrel. 10, Created)  
 DI 01-MAY-1999 (TRENBLrel. 10, Last sequence update)  
 DE 01-NOV-1999 (TRENBLrel. 12, Last annotation update)  
 DE BETA-N-ACETYLGLUCOSAMINIDASE.  
 GN CHIO.  
 OS Pseudalteromonas sp. S9.  
 CC Bacteria; Proteobacteria; gam. subdivision; Alteromonadales.  
 CC Pseudalteromonas.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-S9.  
 RA TEKHARNJANARUK S., GOODEMAN A.S.:  
 RI "Cloning, sequence analysis and characterization of genes involved in  
 RI chitin degradation of a marine bacterium, Pseudalteromonas sp. strain  
 RI S9.";  
 RL Microbiology 0:0-(1999).  
 DR EMBL: AF072374; AAC83237.1;  
 DR HSS: P06865; IQBC.  
 SQ SEQUENCE 78: AA: 88856 MW: FF87EDF8 CRC32:

Query Match 80.7% Score 46: DB 2: Length 78;  
 Best Local Similarity 75.0% Pred. NO. 5.90e-00;  
 Matches 6: Conservative 1: Mismatches 1: Indels 0: Gaps 0:

DB 392 IKRVSNVI 78  
 |||||  
 QY 2 IKRVSNVI 9

RESULT 6 PRELIMINARY: PRI: 192 AA.  
 ID O67344  
 AC O67344  
 DT 01-AUG-1998 (TRENBLrel. 07, Created)  
 DI 01-AUG-1998 (TRENBLrel. 07, Last sequence update)  
 DE 01-NOV-1998 (TRENBLrel. 08, Last annotation update)  
 DE HYPOTHETICAL 22.1 KD PROTEIN.  
 GN AQ1324.  
 OS Aquifex aeolicus.  
 CC Bacteria; Aquificales; Aquificaceae; Aquifex.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-VF5.  
 RX MEDLINE: 98196666.  
 RA DECKERT G., WARREN P.V., GAASTERLAND T., YOUNG W.G., LENOX A.L.,  
 RA GRAHAM D.E., OVERBEK R., SNEAD M.A., KELLER M., AUJAY M., HUBER R.,  
 RA FELDMAN R.A., OVERBEK R., OLSON J.M., SWANSON R.V.:  
 RI "The complete genome of the hyperthermophilic bacterium Aquifex  
 RI aeolicus.";

RL Nature 392:353-358(1998).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN-VF5:

RA DECKER G., WARREN P.V., GAASTERLAND T., YOUNG W.G., LENOX A.L.,

RA GRAHAM D.E., OVERBECK R., SNEAD M.A., KELLER M., AUJAY M., HUBER R.,

RA FEEDMAN R.A., SHORI J.M., OLSON G.J., SWANSON R.V.:

RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.

DR EMBL: AE007341; AAC07306.1; -

KW Hypothetical protein.

SQ SEQUENCE 192 AA: 7528ADDD CRC32:

Query Match 78.9% Score 45: DB 2: Length 192:

Best Local Similarity 66.7% Pred. No. 1.03e+01:

Matches 6: Conservative 2: Mismatches 1: Indels 0: Gaps 0:

Db 111 AIKRVKNNV 119

QY 1 SIKRVSNVI 9

RESULT 7

ID Q9X546 PRELIMINARY: PRI: 293 AA.

AC Q9X546:

DT 01-MAY-1999 (TRENBLrel. 12, Created)

DT 01-NOV-1999 (TRENBLrel. 12, Last sequence update)

DT 01-NOV-1999 (TRENBLrel. 12, Last annotation update)

DE TEICRYCICLINE REPRESSOR PROTEIN HOMOLOG TRH.

GN TRH.

OS Corynebacterium glutamicum.

OC Plasmid R-plasmid PAG2.

CC Bacteria: Firmicutes: Actinobacteria: Actinobacteridae:

CC Actinomycetales: Corynebacterineae: Corynebacteriaceae:

CC Corynebacterium.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-22243:

RA TAUCH A., PFEHLER A., KALINOWSKI J.:

RT "DNA sequence and genetic organization of PAG2, a 19.8-kb R-plasmid of

RT Corynebacterium glutamicum encoding a new class of tetracycline efflux

RT and repressor proteins."

RL Submitted (JAN-1999) to the EMBL/GenBank/ DBJ databases.

DR EMBL: AF121000; AAB25055.1; -

KW Cyclicin: Plasmid.

SQ SEQUENCE 293 AA: 32350 MW: 5FLURAB3 CRC32:

Db 100 SMRRVANVV 108

QY 1 SIKRVSNVI 9

RESULT 8

ID Q9Z3L9 PRELIMINARY: PRI: 360 AA.

AC Q9Z3L9:

DT 01-MAY-1999 (TRENBLrel. 10, Created)

DT 01-MAY-1999 (TRENBLrel. 10, Last sequence update)

DT 01-MAY-1999 (TRENBLrel. 10, Last annotation update)

DE HYPOTHETICAL 41.5 KD PROTEIN.

OS Leptospira borgpetersenii.

CC Bacteria: Spirochaetales: Leptospiraceae: Leptospira.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-L177:

RA KALANBAHET T., BULACH D.M., RAJAKUMAR K., ADLER B.:

RT "Genetic Organization of the Lipopolysaccharide O-antigen Biosynthetic

RT Locus of Leptospira borgpetersenii Serovar Hardjibovis."

RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.

DR EMBL: AF078135; AAD12950.1; -

KW Hypothetical protein.

SQ SEQUENCE 360 AA: 41546 MW: 922C1EE3 CRC32:

Query Match 78.9% Score 45: DB 2: Length 360:

Best Local Similarity 75.0% Pred. No. 1.03e+01:

Matches 6: Conservative Mismatches 1: Indels 0: Gaps 0:

Db 129 IKRVSGVI 136

QY 2 IKRVSNVI 9

RESULT 9

ID C29266 PRELIMINARY: PRI: 569 AA.

AC C29266:

DT 01-JAN-1998 (TRENBLrel. 05, Created)

DT 01-JAN-1998 (TRENBLrel. 05, Last sequence update)

DT 01-NOV-1998 (TRENBLrel. 08, Last annotation update)

DE TYPE 11 SECRETION SYSTEM PROTEIN (CSPE-3).

GN AF0596.

OS Archaeoglobus fulgidus.

CC Archaea: Euryarchaeota: Archaeoglobales: Archaeoglobaceae:

CC Archaeoglobus.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-VC-15 / DSM 4104 / ATCC 49559:

RX MEDLINE: 98C49343.

RA KLENK H.-P., CLAYTON R.A., TOMB J.-F., WHITE G., NELSON K.E.,

RA KETCHUM K.A., DODSON R.J., GWINN M., HICKEY E.K., PETERSON J.D.,

RA RICHARDSON D.L., KERJAVAGE A.R., GRAHAM D.E., KYRPIDES N.C.,

RA FLEISCHMANN R.D., QUACKENBUSH J., LEE N.H., SUTTON G.G., GILL S.,

RA KIRKNESS E.F., DOUGHERTY B.A., MCKENNEY K., ADAMS M.D., LORTUS B.,

RA PETERSON S., REICH C.L., MCNEIL L.K., BADGER J.H., GLODEK A., ZHANG L.,

RA OVERBECK R., GOCCAYNE J.D., WEIDMAN J.F., McDONALD L., UFFERACK T.,

RA COTTON M.D., SPRIGGS T., ARIACH P., KATNE B.P., SYKES S.M.,

RA SADOW P.W., D'ANDREA K.P., HOWMAN C., FUJII C., GARLAND S.A.,

RA MASON T.M., OLSEN G.J., FRASER C.M., SMITH H.G., WOESE C.R.,

RA VENTER J.C.:

RT "The complete genome sequence of the hyperthermophilic, sulphate-

RT reducing archaeon Archaeoglobus fulgidus."

RL Nature 390:364-370(1997).

DR EMBL: AF001035; AAB90245.1; -

DR TIGR: AF0996.

DR PFAM: PF00437; GSPIL\_E.1.

KW Hypothetical protein.

SQ SEQUENCE 569 AA: 56520 MW: 6BF577EB CRC32:

Db 544 IKRVSNVI 551

QY 2 IKRVSNVI 9

RESULT 10

ID O13331 PRELIMINARY: PRI: 682 AA.

AC O13331:

DT 01-JAN-1998 (TRENBLrel. 05, Created)

DT 01-JAN-1998 (TRENBLrel. 05, Last sequence update)

DT 01-NOV-1999 (TRENBLrel. 12, Last annotation update)

DE ACETOLACTATE SYNTHASE.

OS Magnaporthe grisea (Rice blast fungus).

CC Eukaryota: Fungi: Ascomycota: Euascomycetes: Pyrenomycetes:

CC Phylachorales: Magnaporthaceae: Magnaporthae.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-4091-5-8:

RA SWEIGARD J.A., CHUMLEY F.C., ROLL A.M., FARRALL L., VALENT B.:

RL Fungal Genetics Newsletter 44:253(1997).

CC -!- SIMILARITY: WITH OTHER ENZYMES WHICH REQUIRE TPP.

DR EMBL: AF013601; AAB81248.1; -

DR PROSITE: PS00187; TPP\_ENZYMES: 1.

DR PFAM: PF00205: TPP\_enzymes: 1;  
 KW Flavoproteinase: Lyase: Thiamine pyrophosphatase;  
 SQ SEQUENCE 582 AA: 73949 MW: 82F2886: CRC32:

Query Match 78.9% Score 45: DB 3: Length 682;  
 Best Local Similarity 55.6% Pred. No. 1:03e-01;  
 Matches 5: Conservative 4: Mismatches 0: Indels 0: Gaps 0:

DB 279 SIKRVADV 287

1111111111

QY 1 SIKRVSNV 5

RESULT 11  
 ID Q01154 PRELIMINARY: PRT: 582 AA:  
 AC Q01154:  
 DT 01-NOV-1996 (TrEMBLrel: 01, Created)  
 DT 01-NOV-1996 (TrEMBLrel: 01, Last sequence update)  
 DT 01-NOV-1999 (TrEMBLrel: 12, Last annotation update)  
 DE ACETYLACETATE SYNTHASE;  
 OS Magnaporthe oryzae (Rice blast fungus).  
 CC Eukaryota; Fungi; Ascomycota; Euscomycetes; Pyrenomycetes;  
 CC Phyllostales; Magnaporthaceae; Magnaporthe.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-6043:  
 RA SWEIGARD J.A., FARRALL L., VALENT B.S.;  
 RL Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.  
 CC 1: SIMILARITY WITH OTHER ENZYMES WHICH REQUIRE TPP.  
 DR EMBL: U39261; AA65169.1;  
 DR PROSITE: PS0187; TPP\_ENZYMES: 1;  
 DR PFAM: PF00205: TPP\_ENZYMES: 1;  
 KW Flavoproteinase: Lyase: Thiamine pyrophosphatase;  
 SQ SEQUENCE 582 AA: 73873 MW: B5A3D8 CRC32:

Query Match 78.9% Score 45: DB 3: Length 682;  
 Best Local Similarity 55.6% Pred. No. 1:03e-01;  
 Matches 5: Conservative 4: Mismatches 0: Indels 0: Gaps 0:

DB 279 SIKRVADV 287

1111111111

QY 1 SIKRVSNV 5

RESULT 12  
 ID Q65260 PRELIMINARY: PRT: 175 AA:  
 AC Q65260:  
 DT 01-AUG-1998 (TrEMBLrel: 07, Created)  
 DT 01-AUG-1998 (TrEMBLrel: 07, Last sequence update)  
 DT 01-NOV-1999 (TrEMBLrel: 12, Last annotation update)  
 DE P62212.2 PROTEIN  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 CC eumhyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;  
 CC core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;  
 CC Arabidopsis.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CV, COLUMBIA;  
 RA WASHU;  
 RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CV, COLUMBIA;  
 RA GEISEL, C.;  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CV, COLUMBIA;  
 RA WATERSTON R.;  
 RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF058913; AAC13612.1; "

DR PFAM: PF00027: CNMP\_binding: 1;  
 SQ SEQUENCE 175 AA: 19898 MW: 478978F9 CRC32:

Query Match 77.2% Score 44: DB 10: Length 175;  
 Best Local Similarity 55.6% Pred. No. 1:80e-01;  
 Matches 5: Conservative 4: Mismatches 0: Indels 0: Gaps 0:

DB 23 SIKRISGV 21

1111111111

QY 1 SIKRVSNV 9

RESULT 13  
 ID Q67471 PRELIMINARY: PRT: 256 AA:  
 AC Q67471:  
 DT 01-AUG-1998 (TrEMBLrel: 07, Created)  
 DT 01-AUG-1998 (TrEMBLrel: 07, Last sequence update)  
 DT 01-NOV-1999 (TrEMBLrel: 08, Last annotation update)  
 DE HYPOTHELICAL 29.9 KD PROTEIN  
 GN AQ1501.  
 OS Aquifex aeolicus.  
 CC Bacteria; Aquificales; Aquificales; Aquifex  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-VF5:  
 RA DECKERT G., WARREN P.V., GAATKENDT J., YOUNG W.G., LENOX A.L.,  
 RA GRAHAM D.E., OVERBECK R., SNEY M.A., KELLER M., ADAY M., HUBER P.,  
 RA FEDKAN R.A., SHORT J.M., GILSON G., SWANSON R.V.;  
 RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ Databases.  
 DR EMBL: AE00743; AAC07436.1;  
 KW Hypothetical protein.  
 SQ SEQUENCE 256 AA: 29913 MW: F0588B0A CRC32:

Query Match 77.2% Score 44: DB 2: Length 256;  
 Best Local Similarity 44.4% Pred. No. 1:80e-01;  
 Matches 4: Conservative 5: Mismatches 0: Indels 0: Gaps 0:

DB 28 SVKPTGEV 36

1111111111

QY 1 SIKRVSNV 9

RESULT 14  
 ID Q62V54 PRELIMINARY: PRT: 413 AA:  
 AC Q62V54:  
 DT 01-MAY-1999 (TrEMBLrel: 10, Created)  
 DT 01-MAY-1999 (TrEMBLrel: 10, Last sequence update)  
 DT 01-NOV-1999 (TrEMBLrel: 12, Last annotation update)  
 DE F15K9.17 PROTEIN.  
 GN F15K9.17.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 CC eumhyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;  
 CC core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;  
 CC Arabidopsis.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CV, COLUMBIA;  
 RA VISOISKAITA V.S., SCHWARTZ J.R., TORUMI M., YU G., LI J., LIU S.,  
 RA KREMENTSKAYA I., LUROS J., ARAUJO R., BUEHLER E., CONWAY A.B.,  
 RA DEWER K., FENG J., KIM C., LI Y., SHINN P., SUN H., DAVIS R.W.,  
 RA ECKER J.R., FEDERSPIEL N.A., THEOLOGIS A.;  
 RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF058913; AAC13612.1; "



RI Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.  
 RN 2.  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CV, COLUMBIA;  
 RA THEOLOGIS A.;  
 PL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.  
 RN 3.  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CV, COLUMBIA;  
 RA THEOLOGIS A.;  
 PL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.  
 RN 4.  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CV, COLUMBIA;  
 RA THEOLOGIS;  
 KL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AC005278; AAC72119.1; ..  
 DR MENDEL: 33845; Arath:3000:33845.  
 SQ SEQUENCE 433 AA: 45717 MW: 9331822E CRC32:

Query Match 77.2% Score 44 DB 10: Length 433;  
 Best Local Similarity 75.0% Pred. No. 1.80e+01;  
 Matches 6: Conservative 2: Mismatches 0: Indels 0: Gaps 0:

DL 317 SIKRVASV 324  
 I I I I I I  
 QY : SIKRVSNV 8

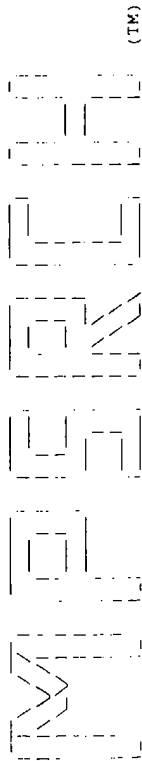
RESULT 15  
 TO QZVSS PRELIMINARY: PRT: 434 AA.  
 AC QZVSS:  
 DT 01-MAY-1999 (TEMBLrel. 10, Created)  
 DT 01-MAY-1999 (TEMBLrel. 10, Last sequence update)  
 DT 01-NOV-1999 (TEMBLrel. 12, Last annotation update)  
 DE F5K9.16 PROTEIN.  
 UN F5K9.16.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 CC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;  
 CC core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;  
 CC Arabidopsis.  
 RN 1.  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CV, COLUMBIA;  
 RA VYSOTSKAYA V.S., SCHWARTZ J.R., TORIUMI M., YU G., LI J., LIU S.,  
 RA KREMERITSKAYA I., LUDOS J., ARAUJO R., BUEHLER E., CONWAY A.B.,  
 RA DEMER K., FENG J., KIM C., LI Y., SHINN P., SUN H., DAVIS R.W.,  
 RA ECKER J.R., FEDERSPEL N.A., THEOLOGIS A.;  
 RI \*Arabidopsis thaliana chromosome 1: BAC F5K9 sequence.\*  
 PL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.  
 RN 2.  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CV, COLUMBIA;  
 RA THEOLOGIS A.;  
 RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.  
 RN 3.  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CV, COLUMBIA;  
 RA THEOLOGIS A.;  
 RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.  
 RN 4.  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CV, COLUMBIA;  
 RA THEOLOGIS;  
 RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AC005278; AAC72120.1; ..  
 DR MENDEL: 33844; Arath:3000:33844.  
 SQ SEQUENCE 434 AA: 46148 MW: 5ED90262 CRC32:

Query Match 77.2% Score 44 DB 10: Length 434;  
 Best Local Similarity 75.0% Pred. No. 1.80e+01;  
 Matches 6: Conservative 2: Mismatches 0: Indels 0: Gaps 0:

Db 318 SIKRVASV 325  
 I I I I I I  
 QY 1 SIKRVSNV 8

Search completed: Mon Jun 19 16:18:10 2000  
 Job time: 10 secs.

\*\*\*\*\*



\*\*\*\*\*

Release 3.1A John F. Collins, BioComputing Research Unit,  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MParch\_pp protein - protein database search, using Smith-Waterman algorithm

Run On: Mon Jun 19 16:11:57 2000 MasPar time 3.71 Seconds  
57,500 Million cell updates/sec

Tabular output not generated.

Title: >US-09-142-524A-14  
Description: (1-9) from US09142524A pep  
Perfect Score: 57  
Sequence: 1 SIKRVSNVI 9

Scoring table: PAM 150  
Gap 15

Searched: 148963 seqs, 23686106 residues

Post-processing: Minimum Watch 0%  
Listing first 45 summaries  
Maximum DB seq length 9

Database: arcdnsseq35  
:cdnseseq

Statistics: Mean 15.426; Variance 44.241; scale 0.349

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

| Result | No. | Score | Match | Length | DB ID | Description                     | Pred. No. |
|--------|-----|-------|-------|--------|-------|---------------------------------|-----------|
| 1      | 53  | 41.0  | 9     | 1      | W5751 | Residues 109-117 of Cr 1.87e-01 |           |

Note: Post-processor removed 44 summaries from list due to search parameters chosen.

# ALIGNMENTS

RESULT 1  
ID W5751 standard: peptide: 9 AA.  
AC W5751:  
DE 17-SEP-1998 (first entry)  
DE Residues 109-117 of Cr 1.  
KW Cr 1 Japanese cedar pollen antigen; allergy: immunotherapy;  
KW HLA class II molecule.  
OS Cryptomeria japonica.  
PN WO9823902-A1  
PC 22-MAY-1998.  
FE 12-NOV-1997; J04129.  
PR 13-NOV-1996; JP-302053.  
PA (MEIP) MELIJI MILK PROD CO LTD.  
FI Dairiki K. Kuro K. Kure A. Sone T.  
DR MPI: 98-295617/25.

PT Peptides derived from Japanese cedar pollen antigens are  
PT immunotherapeutic agents - useful for allergy treatment and typing  
PT HLA class II molecules in allergy sufferers  
PS Claim 12: Page 26: 50pp: Japanese.  
CC This sequence represents residues 109-117 of the Cr 1 protein, and  
CC is a peptide of the invention. The peptides are derived from Japanese  
CC cedar pollen antigens, and are used as immunotherapeutic agents in the  
CC treatment of allergy. The peptides can be used for identification and  
CC typing of the particular HLA class II molecules in an allergy sufferer,  
CC and also for peptide immunotherapy of an allergy. Using these peptides  
CC the immunotherapy can be targeted more specifically to the requirements  
CC of the individual patient, allowing more effective treatment of an  
CC allergy, including those patients for whom treatment with a conventional  
CC immunotherapeutic agent is ineffective.  
SQ Sequence 9 AA:

Query Match 93.0%; Score 53; DB ID Length 9;  
Best Local Similarity 100.0%; Pred. No. 1.87e-01;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 2 :KRVSNVI: 9  
QY 2 :KRVSNVI: 9

Search completed: Mon Jun 19 16:12:04 2000  
Job time : 7 secs.

.....  
XREF (TM)  
.....

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MFarch\_PP protein - protein database search, using Smith-Waterman algorithm

Run on: Tue Jun 20 13:39:35 2000 MasPar time 2.16 Seconds  
60.035 Million cell updates/sec

Tabular output not generated.

File: US-09-142-524A-14  
Description: (149) from USC0142524A.fpp  
Perfect Score: 87  
Sequence: 1 SKRVSNI: 9

Scoring table: PAM 150  
Gap 15

Searched: 145141 seqs, 14437480 residues

Post-processing: Minimum Match 04  
Listing first 45 summaries  
Maximum DB seq length 9

Database: a-issued  
1:5A\_COMB 2:5B\_COMB 3:6\_COM 4:1PT\_COMB 5:backfiles1

Statistics: Mean 14.393, Variance 44.418, scale 0.324

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result Query  
No. Score Match Length DB ID Description Pred. No.  
.....

Summary

No matches found

Search completed: Tue Jun 20 13:39:38 2000  
Job time : 3 secs.

\*\*\*\*\*  
 WIRELESS (TM)  
 \*\*\*\*\*

Release 3.1A John F. Collins, BioComputing Research Unit.  
 Copyright (c) 1993-1998 University of Edinburgh, U.K.  
 Distribution rights by Oxford Molecular Ltd

MSrch\_off protein - protein database search, using Smith-Waterman algorithm  
 Run on: Mon Jun 19 16:12:21 2000; MasPar time 14.17 Seconds  
 63.988 Million cell updates/sec  
 Tabular output not generated.

Title: US-09-142-524A-14  
 Description: (1-9) from US09142524A.pep  
 Perfect Score: 5  
 Sequence: 1 S:KRVSNV: 9  
 Scoring table: PAM 150  
 Gap 15

Searched: 721208 seqs, 100765575 residues  
 Post-processing: Minimum Match 0%  
 Listing first 45 summaries  
 Maximum DB seq length 9

Database: a-pending  
 1:PCI 2:UG 3:UG 4:UG 5:UG 6:UG 7:UG 8:UG 9:UG  
 10:UG 11:UG 12:UG 13:UG 14:UG 15:UG 16:UG 17:UG  
 18:UG 19:UG 20:UG 21:UG 22:UG 23:UG 24:UG  
 25:NEWC9

Statistics: Mean 17.247; Variance 40.530; Scale 0.426  
 Pred. No. is the number of results predicted by chance to have a  
 score greater than or equal to the score of the result being printed.  
 and is derived by analysis of the total score distribution.

SUMMARIES  

| Result No.       | Query | Score | Match | Length | DB ID | Description | Pred. No. |
|------------------|-------|-------|-------|--------|-------|-------------|-----------|
| No matches found |       |       |       |        |       |             |           |

Search completed: Mon Jun 19 16:12:44 2000  
 Job time : 23 secs.

\*\*\*\*\*



(TM)

\*\*\*\*\*

Release 1.1A John F. Collins, BioComputing Research Unit,  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_PP protein - protein database search, using Smith-Waterman algorithm

Run on: Mon Jun 19 16:11:24 2000 Maspar time 3.89 Seconds  
109.175 Million cell updates/sec

Tabular output not generated.

Title: US-09-142-524A-14  
Description: (1-9) from US09142524A.pep  
Perfect Score: 57  
Sequence: : SIKKVSNI: 9

Scoring table: FAW 150  
Gap 15

Searched: 142080 seqs, 47172406 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries  
Maximum DB seq length 9

Database: pir62  
pir61 2:pir2 3:pir3 4:pir4

Statistics: Mean 22.352; Variance 25.242; scale 0.886

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result | Query | No. | Score | Match | Length | DB | ID | Description | Pred. No. |
|--------|-------|-----|-------|-------|--------|----|----|-------------|-----------|
|--------|-------|-----|-------|-------|--------|----|----|-------------|-----------|

No matches found.

Search completed: Mon Jun 19 16:11:40 2000  
Job time 1:16 secs.

\*\*\*\*\*

The logo consists of the word "WELLS" in a large, bold, sans-serif font. To the right of "WELLS" is the word "(TM)" in a smaller font. The entire logo is enclosed within a rectangular border.

\*\*\*\*\*

Release 3.1A John F. Collins, BioComputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MSEARCH\_PEP protein - protein database search, using Smith-Waterman algorithm

Run on: Mon Jun 19 16:10:32 2000: MasPar time 2.50 seconds  
109.525 Million cell updates/sec

Tabular output not generated.

Title: >US-09-142-524A-14  
Description: (1-9); from US09142524A.pep  
Perfect Score: 57  
Sequence: 1 S:KRVSNV1 9

Scoring table: PAM 150  
Gap 15

Searched: 43857 seqs, 30454973 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries  
Maximum DB seq length 9

Database: swiss-prot38  
1:swissprot

Statistics: Mean 22.858; Variance 21.701; scale 1.053

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

| Result | Query | %                  |             |           |
|--------|-------|--------------------|-------------|-----------|
| No.    | Score | Match Length (H L) | Description | Pred. No. |

No matches found.

Search completed: Mon Jun 19 16:10:32 2000  
Job time : 12 secs.

.....  
M E S E R F A  
..... (TM)  
.....

Release 3.1A John F. Collins, Biocomputing Research Unit,  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

Mpsrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Mon Jun 19 16:10:49 2000; Maspar time 5.97 Seconds  
104.457 Million cell updates/sec

Tabular output not generated.

Title: US-09-142-524A-14  
Description: (1-9) from US09142524A.ppt  
Perfect Score: 57  
Sequence: 1 SKKRVSNV: 9

Scoring table: PAM 150  
Gap 15

Searched: 225878 seqs, 69334122 residues

Post-processing: Minimum Match: 0%  
Listing first 45 summaries  
Maximum DB seq length 9

Database: strembii2  
1:sp-archaea 2:sp-bacteria 3:sp-fungi 4:sp-human  
5:sp-invertebrate 6:sp-mammal 7:sp-the 8:sp-organelle  
9:sp-phage 10:sp-plant 11:sp-rodent 12:sp-unclassified  
13:sp-vertebrate 14:sp-virus

Statistics: Mean 21.680; Variance 21.726; scale 1.007

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result<br>No. | Query<br>Score | Match Length | DB ID | Description | Pred. No. |
|---------------|----------------|--------------|-------|-------------|-----------|
|---------------|----------------|--------------|-------|-------------|-----------|

NC matches found.

Search completed: Mon Jun 19 16:11:08 2000  
Job time : 19 secs.

\*\*\*\*\*  
 AMBUREL (TM)  
 \*\*\*\*\*

Release 3.2A John F. Collins, BioComputing Research Unit.  
 Copyright: (c) 1993-1998 University of Edinburgh, U.K.  
 Distribution rights by Oxford Molecular Ltd

MSrch\_PP protein - protein database search, using Smith-Waterman algorithm  
 Run on: Mon Jun 19 16:06:50 2000: MasPar time 3.66 Seconds  
 58,202 Million cell updates/sec  
 Tabular output not generated.

Title: US-09-142-524A-7  
 Description: (1-9) from US09142524A.1ep  
 Perfect Score: 64  
 Sequences: 1 FIKRVSNVI: 4

Scoring table: PAM 150  
 Gap 15

Searched: 188963 seqs, 23686106 residues

Post-processing: Minimum Match 0%  
 Listing first 45 summaries

Database: a-geneseq35  
 l-geneseqp

Statistics: Mean 16.765, Variance 46.25, scale 0.362

Pred. No. is the number of results predicted by chance to have a  
 score greater than or equal to the score of the result being printed,  
 and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Query Match | Length | DB ID    | Description             | Pred. No. |
|------------|-------|-------------|--------|----------|-------------------------|-----------|
| 1          | 64    | 100.0       | 9      | 1 W57751 | Residues 109-117 of Cr  | 1.25e+00  |
| 2          | 64    | 100.0       | 15     | 1 W57750 | Residues 106-120 of Cr  | 1.25e+00  |
| 3          | 64    | 100.0       | 15     | 1 R82921 | Japanese cedar pollen   | 1.25e+00  |
| 4          | 64    | 100.0       | 20     | 1 R45552 | Cry j 1 pollen allergen | 1.25e+00  |
| 5          | 64    | 100.0       | 20     | 1 R82501 | Cry j 1 Japanese Cedar  | 1.25e+00  |
| 6          | 64    | 100.0       | 30     | 1 W44683 | T-cell epitope peptide  | 1.25e+00  |
| 7          | 64    | 100.0       | 31     | 1 W27372 | Multi-epitope peptide   | 1.25e+00  |
| 8          | 64    | 100.0       | 31     | 1 W27373 | Multi-epitope peptide   | 1.25e+00  |
| 9          | 64    | 100.0       | 80     | 1 W27369 | Multi-epitope peptide   | 1.25e+00  |
| 10         | 64    | 100.0       | 105    | 1 W27370 | Multi-epitope peptide   | 1.25e+00  |
| 11         | 64    | 100.0       | 134    | 1 W27371 | Multi-epitope peptide   | 1.25e+00  |
| 12         | 64    | 100.0       | 353    | 1 R75388 | Japanese cedar pollen   | 1.25e+00  |
| 13         | 64    | 100.0       | 353    | 1 R81587 | Cedar pollen allergen   | 1.25e+00  |
| 14         | 64    | 100.0       | 374    | 1 R60166 | Japanese cedar pollen   | 1.25e+00  |
| 15         | 64    | 100.0       | 374    | 1 R31537 | Cry j 1                 | 1.25e+00  |
| 16         | 64    | 100.0       | 374    | 1 R82490 | Cry j 1 Japanese Cedar  | 1.25e+00  |
| 17         | 64    | 100.0       | 374    | 1 R45541 | Cry j 1 pollen allergen | 1.25e+00  |
| 18         | 49    | 76.6        | 292    | 1 R04895 | Penicillinase-insulin   | 8.03e-01  |
| 19         | 48    | 75.0        | 367    | 1 R45577 | Jun s 1                 | 1.05e-02  |
| 20         | 48    | 75.0        | 370    | 1 R45578 | Jun v 1                 | 1.05e-02  |
| 21         | 47    | 73.4        | 424    | 1 W86150 | Human growth-related C  | 1.36e-02  |
| 22         | 46    | 71.9        | 20     | 1 R45553 | Cry j 1 pollen allergen | 1.77e-02  |
| 23         | 46    | 71.9        | 20     | 1 R82502 | Cry j 1 Japanese Cedar  | 1.77e-02  |

## ALIGNMENTS

RESULT 1  
 ID W57751 standard: peptide: 9 AA.  
 AC W57751:  
 DI 17-SEP-1998 (first entry)  
 DE Residues 109-117 of Cry j 1  
 KW Cry j 1: Japanese cedar pollen antigen: allergy: immunotherapy:  
 KW HLA class II molecule:  
 OS Cryptomeria japonica:  
 PN W09820902-A1:  
 PD 22-MAY-1998:  
 PF 22-NOV-1997: J04129  
 PR 13-NOV-1996: JP-302053:  
 PA (MEIP) MEIJI MILK PROD CO LTD:  
 PI Dairiki K, Kino K, Kume A, Sone T:  
 DR WPI: 98-297617/26:  
 PT Peptides derived from Japanese cedar pollen antigens are  
 PT immunotherapeutic agents useful for allergy treatment and typing  
 PT HLA class II molecules in allergy sufferers  
 PS Claim 12: Page 24: 50pp: Japanese:  
 CC This sequence represents residues 109-117 of the Cry j 1 protein, and  
 CC is a peptide of the invention. The peptides are derived from Japanese  
 CC cedar pollen antigens, and are used as immunotherapeutic agents in the  
 CC treatment of allergy. The peptides can be used for identification and  
 CC typing of the particular HLA class II molecules in an allergy sufferer,  
 CC and also for peptide immunotherapy of an allergy. Using these peptides,  
 CC the immunotherapy can be targeted more specifically to the requirements  
 CC of the individual patient, allowing more effective treatment of an  
 CC allergy, including those patients for whom treatment with a conventional  
 CC immunotherapeutic agent is ineffective.  
 SQ Sequence 9 AA:  
 Query Match 100.0%; Score 64; DB 1; Length 9;  
 Best Local Similarity 100.0%; Pred. No. 1.25e+00;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Eb 1 FIKRVSNVI 9  
 QY 1 FIKRVSNVI 9  
 RESULT 2  
 ID W57750 standard: peptide: 15 AA.  
 AC W57750:  
 DI 17-SEP-1998 (first entry)  
 DE Residues 106-120 of Cry j 1  
 KW Cry j 1: Japanese cedar pollen antigen: allergy: immunotherapy:  
 KW HLA class II molecule:



Cryptomeria japonica.  
W09820502-A1.  
22-MAY-1996: WO-00139.  
13-NOV-1996: JP-302053.  
(MEIP) MEIJU MILK PROD CO LTD.  
Dairiki K. Kino K. Kume A. Sone T.  
WPI: 98-23767/26.  
Peptides derived from Japanese cedar pollen antigens are immunotherapeutic agents - useful for allergy treatment and typing HLA class II molecules in allergy sufferers.  
Claim 12: Page 26: 5pp: Japanese.  
This sequence represents residues 106-120 of the Cry j 1 protein, and is a peptide of the invention. The peptides are derived from Japanese cedar pollen antigens, and are used as immunotherapeutic agents in the treatment of allergy. The peptides can be used for identification and typing of the particular HLA class II molecules in an allergy sufferer, and also for peptide immunotherapy of an allergy. Using these peptides the immunotherapy can be targeted more specifically to the requirements of the individual patient, allowing more effective treatment of an allergy, including those patients for whom treatment with a conventional immunotherapeutic agent is ineffective.  
Sequence: 15 AA:

Query Match: 100.0%; Score 64; DB 1; Length 15;  
Best Local Similarity 100.0%; Pred. No. 1.25e+00;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

4 FIKRVSNVI 12  
|||||||  
1 FIKRVSNVI 9

R82991 standard; peptide: 15 AA.

12-MAR-1996 (first entry)  
Japanese cedar pollen allergen Cry j 1 derived T-cell epitope peptide.  
Japanese cedar pollen allergen; Cry j 1; T-cell epitope; peptides.  
Prevention; treatment; cryptomeria pollenosis.  
Cryptomeria japonica.  
J0718295-A.  
09-MAY-1995: 262626.  
20-OCT-1993: JP-262626.  
(MEIP) MEIJU MILK PROD CO LTD.  
WPI: 95-20384/27.  
New cryptomeria pollen allergen T cell epitope peptide - used for prevention, treatment and investigation of Japanese cedar pollenosis.  
Claim 5: Page 2: 8pp: Japanese.  
R75388 is the Japanese cedar pollen allergen Cry j 1, from which the T-cell epitope peptides R89289-R89295 were derived. The peptides can be used for the prevention and treatment of cryptomeria pollenosis, and also for the investigation of pollenosis.

Query Match: 100.0%; Score 64; DB 1; Length 15;  
Best Local Similarity 100.0%; Pred. No. 1.25e+00;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

4 FIKRVSNVI 12  
|||||||  
1 FIKRVSNVI 9

R45552 standard; Protein: 20 AA.

13-JUL-1994 (first entry)  
Cry j 1 pollen allergen peptide CJI-11.  
Japanese cedar; detection; allergy; treatment; diagnosis; T cell epitope; sensitivity.  
Cryptomeria japonica.

W09401560-A.  
20-JAN-1994: U00139.  
15-JAN-1993: WO-005861.  
10-JUL-1992: US-938990.  
(IMMU-) IMMULOGIC PHARM CORP.  
Bond JF, Garman RD, Griffith LJ, Kuo M, Pollock Z.  
WPI: 94-035066/04.  
Antigens derived from Japanese cedar pollen allergen; Cry j 1; contain at least two T cell epitope(s). used to treat or diagnose allergy.  
Claim 1: Fig 13: 137pp: English.  
The sequence is that of an isolated peptide of the Japanese cedar pollen allergen Cry j 1 (amino acids 101-120). The peptide, CJI-11, can be used for the treatment and diagnosis of allergies associated with Japanese cedar pollen. It has enhanced therapeutic properties but reduced side effects compared to naturally occurring allergens.  
Sequence: 20 AA:

Query Match: 100.0%; Score 64; DB 1; Length 20;  
Best Local Similarity 100.0%; Pred. No. 1.25e+00;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

9 FIKRVSNVI 17  
|||||||  
1 FIKRVSNVI 9

R82501 standard; Protein: 20 AA.  
15-APR-1996 (first entry)  
Cry j 1 Japanese Cedar pollen allergen; peptide fragment CJI-11.  
Cry j 1; Japanese cedar pollen allergen; modified; drug production; allergy; Cryptomeria japonica.  
Cryptomeria japonica.  
W09527786-A1.  
19-OCT-1995:  
06-APR-1995: U04245.  
08-APR-1994: US-226248.  
06-DEC-1994: US-350225.  
(IMMU-) IMMULOGIC PHARM CORP.  
Chen X, Evans S, Franzen HM, Kuo M, Powers SP;  
Shaked Z;  
WPI: 95-36639/47.  
Modified cryptomeria japonica (Cry j) 1 peptide(s) - useful for treating allergy to Japanese cedar pollen allergen or immunologically cross reactive allergens  
Disclosure: Figure 2: 6pp: English.

Novel peptides of Cry j 1 have been modified as a part of a preformulation scheme to develop an optimised drug product for therapeutic treatment of humans suffering from allergy to Japanese cedar pollen allergen or an allergen which is immunologically cross reactive with Japanese cedar pollen allergen. Such modified peptides possess certain characteristics which render them particularly suitable for drug product formulation. Peptide fragments of Cry j 1, modified and unmodified, are given in R82491-R82525. This peptide fragment corresponds to amino acids 101-120 of the allergen mature protein.  
Sequence: 20 AA:

Query Match: 100.0%; Score 64; DB 1; Length 20;  
Best Local Similarity 100.0%; Pred. No. 1.25e+00;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

9 FIKRVSNVI 17  
|||||||  
1 FIKRVSNVI 9

R82525 standard; peptide: 30 AA.  
W44683  
W44683:

DT 01-MAY-1998 (first entry)  
 DE Multi-epitope peptide used as immunotherapeutic agent #3.  
 KW T-cell epitope peptide; sugi pollen antigen; sugi pollinosis.  
 OS Synthetic.  
 PN Cryptomeria japonica.  
 PC J1002700-A.  
 PD 13-JAN-1998.  
 PF 24-JUN-1996; 163287.  
 PR 24-JUN-1996; JP-080702.  
 PA (DAIICHI) DAIICHI CHEM IND LTD.  
 PI (DAIICHI) DAIICHI SEIKA KAISHA LTD.  
 DR WPI: 98-13363/13.  
 PT T cell epitope peptide of sugi pollen antigen - used in the treatment of sugi pollinosis.  
 PS Claim 1: Page 4; 34pp; Japanese.  
 CC T-cell epitope peptides W41682-88 and their derivatives react with sugi pollinosis patient peripheral blood T lymphocytes. A composition prepared by combining at least 2 of the above peptides and/or their derivatives is used for the prevention and treatment of sugi pollinosis.  
 SQ Sequence 30 AA;  
 Query Match 100.0%; Score 64; DB 1; Length 30;  
 Best Local Similarity 100.0%; Pred. No. 1.25e+00;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 DB 9 FIKRVSNVI 17  
 QY :|||||  
 : FIKRVSNVI 9  
 RESULT 7  
 ID W27372 standard: peptide: 11 AA.  
 AC W27372  
 DT 24-MAR-1998 (first entry)  
 DE Multi-epitope peptide used as immunotherapeutic agent #4.  
 KW Multi-epitope peptide; immunotherapeutic agent; allergic disease.  
 OS Synthetic.  
 PN W09732600-A.  
 PC 12-SEP-1997.  
 PD 10-MAR-1997; J00740.  
 PR 10-MAR-1997; JP-080702.  
 PA (MEIP) MEIJI MILK PROD CO LTD.  
 PI Daiichi K. Iwama A. Kino K. Kume A. Some T.  
 DR WPI: 97-470495/43.  
 PT Peptide immunotherapeutic agent to treat allergic diseases - contains multi-epitope peptide containing T cell epitope regions from different allergens.  
 PS Claim 9: Page 32; 58pp; Japanese.  
 CC The present sequence represents a multi-epitope peptide which is used as a new immunotherapeutic agent. It comprises T cell epitope regions from 2 or more different allergens (preferably linked via arginine or lysine dimers), where the T cell epitope regions have a positivity index greater than 100 as measured in a patient group responding to the allergen; have at least 70% reactivity with lymphocytes from patients responding to the allergen; and are not reactive with immunoglobulin E (IgE) antibodies from patients responsive to the allergen. The agent can be used to prevent and treat a wide variety of allergic diseases, e.g. by desensitisation. Side effects, e.g. those mediated by IgE, are reduced.  
 SQ Sequence 31 AA;  
 Query Match 100.0%; Score 64; DB 1; Length 31;  
 Best Local Similarity 100.0%; Pred. No. 1.25e+00;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 DB 23 FIKRVSNVI 31  
 QY :|||||  
 : FIKRVSNVI 9  
 RESULT 8  
 ID W27373 standard: peptide: 11 AA.  
 AC W27373  
 DT 24-MAR-1998 (first entry)  
 DE Multi-epitope peptide used as immunotherapeutic agent #1.  
 KW Multi-epitope peptide; immunotherapeutic agent; allergic disease.  
 OS Synthetic.  
 PN W09732600-A.  
 PC 12-SEP-1997.  
 PD 10-MAR-1997; J00740.  
 PR 10-MAR-1997; JP-0807-2.  
 PA (MEIP) MEIJI MILK PROD CO LTD.  
 PI Daiichi K. Iwama A. Kino K. Kume A. Some T.  
 DR WPI: 97-470495/43.  
 PT Peptide immunotherapeutic agent to treat allergic diseases - contains multi-epitope peptide containing T cell epitope regions from different allergens.  
 PS Claim 9: Page 32; 58pp; Japanese.  
 CC The present sequence represents a multi-epitope peptide which is used as a new immunotherapeutic agent. It comprises T cell epitope regions from 2 or more different allergens (preferably linked via arginine or lysine dimers), where the T cell epitope regions have a positivity index greater than 100 as measured in a patient group responding to the allergen; have at least 70% reactivity with lymphocytes from patients responding to the allergen; and are not reactive with immunoglobulin E (IgE) antibodies from patients responsive to the allergen. The agent can be used to prevent and treat a wide variety of allergic diseases, e.g. by desensitisation. Side effects, e.g. those mediated by IgE, are reduced.  
 SQ Sequence 31 AA;  
 Query Match 100.0%; Score 64; DB 1; Length 31;  
 Best Local Similarity 100.0%; Pred. No. 1.25e+00;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 DB 23 FIKRVSNVI 31  
 QY :|||||  
 : FIKRVSNVI 9  
 RESULT 9  
 ID W27369 standard: peptide: 11 AA.  
 AC W27369  
 DT 24-MAR-1998 (first entry)  
 DE Multi-epitope peptide used as immunotherapeutic agent #1.  
 KW Multi-epitope peptide; immunotherapeutic agent; allergic disease.  
 OS Synthetic.  
 PN W09732600-A.  
 PC 12-SEP-1997.  
 PD 10-MAR-1997; J00740.  
 PR 10-MAR-1997; JP-0807-2.  
 PA (MEIP) MEIJI MILK PROD CO LTD.  
 PI Daiichi K. Iwama A. Kino K. Kume A. Some T.  
 DR WPI: 97-470495/43.  
 PT Peptide immunotherapeutic agent to treat allergic diseases - contains multi-epitope peptide containing T cell epitope regions from different allergens.  
 PS Claim 6: Page 31; 58pp; Japanese.  
 CC The present sequence represents a multi-epitope peptide which is used as a new immunotherapeutic agent. It comprises T cell epitope regions from 2 or more different allergens (preferably linked via arginine or lysine dimers), where the T cell epitope regions have a positivity index greater than 100 as measured in a patient group responding to the allergen; have at least 70% reactivity with lymphocytes from patients responding to the allergen; and are not reactive with immunoglobulin E (IgE) antibodies from patients responsive to the allergen. The agent can be used to prevent and treat a wide variety of allergic diseases, e.g. by desensitisation. Side effects, e.g. those mediated by IgE, are reduced.  
 SQ Sequence 80 AA;  
 Query Match 100.0%; Score 64; DB 1; Length 80;  
 Best Local Similarity 100.0%; Pred. No. 1.25e+00;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 DB 17 FIKRVSNVI 25

DT 24-MAR-1998 (first entry)  
 DE Multi-epitope peptide used as immunotherapeutic agent #3.  
 KW Multi-epitope peptide; immunotherapeutic agent; allergic disease.  
 OS Synthetic.  
 PN W09732600-A.  
 PC 12-SEP-1997.  
 PD 10-MAR-1997; J00740.  
 PR 10-MAR-1997; JP-080702.  
 PA (MEIP) MEIJI MILK PROD CO LTD.  
 PI Daiichi K. Iwama A. Kino K. Kume A. Some T.  
 DR WPI: 97-470495/43.  
 PT Peptide immunotherapeutic agent to treat allergic diseases - contains multi-epitope peptide containing T cell epitope regions from different allergens.  
 PS Claim 9: Page 32; 58pp; Japanese.  
 CC The present sequence represents a multi-epitope peptide which is used as a new immunotherapeutic agent. It comprises T cell epitope regions from 2 or more different allergens (preferably linked via arginine or lysine dimers), where the T cell epitope regions have a positivity index greater than 100 as measured in a patient group responding to the allergen; have at least 70% reactivity with lymphocytes from patients responding to the allergen; and are not reactive with immunoglobulin E (IgE) antibodies from patients responsive to the allergen. The agent can be used to prevent and treat a wide variety of allergic diseases, e.g. by desensitisation. Side effects, e.g. those mediated by IgE, are reduced.  
 SQ Sequence 31 AA;  
 Query Match 100.0%; Score 64; DB 1; Length 31;  
 Best Local Similarity 100.0%; Pred. No. 1.25e+00;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 DB 23 FIKRVSNVI 31  
 QY :|||||  
 : FIKRVSNVI 9  
 RESULT 9  
 ID W27369 standard: peptide: 11 AA.  
 AC W27369  
 DT 24-MAR-1998 (first entry)  
 DE Multi-epitope peptide used as immunotherapeutic agent #1.  
 KW Multi-epitope peptide; immunotherapeutic agent; allergic disease.  
 OS Synthetic.  
 PN W09732600-A.  
 PC 12-SEP-1997.  
 PD 10-MAR-1997; J00740.  
 PR 10-MAR-1997; JP-0807-2.  
 PA (MEIP) MEIJI MILK PROD CO LTD.  
 PI Daiichi K. Iwama A. Kino K. Kume A. Some T.  
 DR WPI: 97-470495/43.  
 PT Peptide immunotherapeutic agent to treat allergic diseases - contains multi-epitope peptide containing T cell epitope regions from different allergens.  
 PS Claim 6: Page 31; 58pp; Japanese.  
 CC The present sequence represents a multi-epitope peptide which is used as a new immunotherapeutic agent. It comprises T cell epitope regions from 2 or more different allergens (preferably linked via arginine or lysine dimers), where the T cell epitope regions have a positivity index greater than 100 as measured in a patient group responding to the allergen; have at least 70% reactivity with lymphocytes from patients responding to the allergen; and are not reactive with immunoglobulin E (IgE) antibodies from patients responsive to the allergen. The agent can be used to prevent and treat a wide variety of allergic diseases, e.g. by desensitisation. Side effects, e.g. those mediated by IgE, are reduced.  
 SQ Sequence 80 AA;  
 Query Match 100.0%; Score 64; DB 1; Length 80;  
 Best Local Similarity 100.0%; Pred. No. 1.25e+00;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 DB 17 FIKRVSNVI 25

```

1111111
1 FIKRVSNVI 9

RESULT 10
ID W27370 standard; peptide: 105 AA.
AC W27370:
DE 24-MAR-1998 (first entry)
KW Multi-epitope peptide used as immunotherapeutic agent #2.
KW Multi-epitope peptide: immunotherapeutic agent; allergic disease.
KW T-cell epitope region: allergen; lymphocyte; immunoglobulin E.
OS Synthetic.
PN W09732600-A1.
PD 12-SEP-1997.
PR 10-MAR-1997: J00740.
PR 10-MAR-1996: JP-080702.
PA (MEIP ) MEIJI MILK PROD CO LTD.
PA Dairiki K, Iwama A, Kino K, Kume A, Sone T;
WP1: 97-470495/43.
PT Peptide immunotherapeutic agent to treat allergic diseases -
PT contains multi-epitope peptide containing T cell epitope regions
PT from different allergens
PS Claim 5: Page 31: 58pp; Japanese.
CC The present sequence represents a multi-epitope peptide which is used as
CC a new immunotherapeutic agent. It comprises T cell epitope regions from 2
CC or more different allergens (preferably linked via arginine or lysine
CC dimers), where the T cell epitope regions have a positivity index
CC greater than 100 as measured in a patient group responding to the
CC allergen; have at least 70% reactivity with lymphocytes from patients
CC responding to the allergen; and are not reactive with immunoglobulin E
CC (IgE) antibodies from patients responsive to the allergen. The agent can
CC be used to prevent and treat a wide variety of allergic diseases, e.g. by
CC desensitisation. Side effects, e.g. those mediated by IgE, are reduced.
SQ Sequence 105 AA.

Query Match 100.0% Score 64: DB 1: Length 105;
Best Local Similarity 100.0% Pred. No. 1.25e-00;
Matches 9: Conservative 0: Mismatches 0: Indels 0: Gaps 0;

17 FIKRVSNVI: 25
1111111
1 FIKRVSNVI: 9

RESULT 11
ID W27371 standard; peptide: 114 AA.
AC W27371:
DE 24-MAR-1998 (first entry)
KW Multi-epitope peptide used as immunotherapeutic agent #3.
KW Multi-epitope peptide: immunotherapeutic agent; allergic disease.
KW T-cell epitope region: allergen; lymphocyte; immunoglobulin E.
OS Synthetic.
PN W09732600-A1.
PD 12-SEP-1997.
PR 10-MAR-1997: J00740.
PR 10-MAR-1996: JP-080702.
PA (MEIP ) MEIJI MILK PROD CO LTD.
PA Dairiki K, Iwama A, Kino K, Kume A, Sone T;
WP1: 97-470495/43.
PT Peptide immunotherapeutic agent to treat allergic diseases -
PT contains multi-epitope peptide containing T cell epitope regions
PT from different allergens
PS Claim 5: Page 32: 58pp; Japanese.
CC The present sequence represents a multi-epitope peptide which is used as
CC a new immunotherapeutic agent. It comprises T cell epitope regions from 2
CC or more different allergens (preferably linked via arginine or lysine
CC dimers), where the T cell epitope regions have a positivity index
CC greater than 100 as measured in a patient group responding to the
CC allergen; have at least 70% reactivity with lymphocytes from patients
CC responding to the allergen; and are not reactive with immunoglobulin E
CC (IgE) antibodies from patients responsive to the allergen. The agent can
CC be used to prevent and treat a wide variety of allergic diseases, e.g. by
CC desensitisation. Side effects, e.g. those mediated by IgE, are reduced.
SQ Sequence 114 AA.

```

```

SQ Sequence 134 AA:

Query Match 100.0% Score 64: DB 1: Length 134;
Best Local Similarity 100.0% Pred. No. 1.25e-00;
Matches 9: Conservative 0: Mismatches 0: Indels 0: Gaps 0;

17 FIKRVSNVI: 25
1111111
1 FIKRVSNVI: 9

RESULT 12
ID R75388 standard; protein: 353 AA.
AC R75388:
DE 12-MAR-1996 (first entry)
DE Japanese cedar pollen allergen Cryj 1.
KW Japanese cedar; pollen allergen; Cryj 1; T-cell epitope; peptides.
KW prevention; treatment; cryptomeria pollinosis.
OS Cryptomeria japonica.
PH Key: Location/Qualifiers
FT peptide 51..75
FT /note= "T-cell epitope peptide"
FT peptide 91..105
FT /note= "T-cell epitope peptide"
FT peptide 106..120
FT /note= "T-cell epitope peptide"
FT peptide 146..160
FT /note= "T-cell epitope peptide"
FT peptide 211..225
FT /note= "T-cell epitope peptide"
FT peptide 326..340
FT /note= "T-cell epitope peptide"
FT peptide 335..346
FT /note= "T-cell epitope peptide"
PN J07118295-A.
PD 09-MAY-1995.
PR 20-OCT-1993: 262626.
PR 20-OCT-1993: JP-262626.
PA (MEIP ) MEIJI MILK PROD CO LTD.
PA WP1: 95-203834/27.
PT New cryptomeria pollen allergen T-cell epitope peptide - used for
PT prevention, treatment and investigation of Japanese cedar pollinosis
PT Disclosure; Figs 1-2: 8pp; Japanese.
CC R75388 is the Japanese cedar pollen allergen Cryj 1, from which the
CC T-cell epitope peptides R89249-R89295 were derived. The peptides
CC can be used for the prevention and treatment of cryptomeria
CC pollinosis, and also for the investigation of pollinosis.
SQ Sequence 353 AA:

Query Match 100.0% Score 64: DB 1: Length 453;
Best Local Similarity 100.0% Pred. No. 1.25e-00;
Matches 9: Conservative 0: Mismatches 0: Indels 0: Gaps 0;

109 FIKRVSNVI: 117
1111111
1 FIKRVSNVI: 9

RESULT 13
ID R81587 standard; protein: 353 AA.
AC R81587:
DE 24-MAY-1996 (first entry)
DE Cedar pollen allergen B.
KW Cedar; pollen; allergen; immunoglobulin E; IgE; T-cell epitope;
KW antibody; pollinosis; therapy; immunotherapy.
OS Cryptomeria japonica.
PN EP-700929-A2.
PD 13-MAR-1996.
PR 08-SEP-1995: 306295.
PR 10-SEP-1994: JP-242137.
PR 14-JUL-1995: JP-200221.
PR 14-JUL-1995: JP-200204.
PA (HAYB ) HAYASHIBARA SEIBUTSU KAGAKU.

```

```

PI Hino K., Saito S., Taniguchi Y.
DR WPI: 96-140976/25.
PI New peptide(s) derived from cedar pollen allergens - activate
PI allergen-specific T-cells, but not allergen-specific IgE antibodies.
PI used for treating cedar pollinosis.
PS Claim 5: Page 31-32: 35pp: English
CC Synthetic peptides based on portions of cedar pollen allergens A
CC (R81586) and B (R81587) were tested for their ability to activate
CC cedar allergen-specific T-cells, but not allergen-specific IgE
CC antibodies. 6 Peptides (R81580-R81585) were identified as T-cell
CC epitopes. These peptides, plus subsequences (R81573-79) essential
CC for T-cell recognition, and homologous peptides (R81588-96) can
CC be used as immunotherapeutic agents to treat or prevent cedar
CC pollinosis, avoiding side-effects such as anaphylaxis.
SQ Sequence 35; AA:

Query Match 100.0%; Score 64; DB 1; Length 374;
Best Local Similarity 100.0%; Pred. No. 1,256-00;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DL 139 FIKRVSNVI 117
      11111111
QY : FIKRVSNVI 9

RESULT 14
ID R60166 standard: Protein; 374 AA.
AC R60166;
DI 24-MAR-1995 (first entry)
DE Japanese cedar pollen antigen Cryj1.
KW Japanese cedar; pollen antigen; allergen; Cryj1; sugi; pollinosis.
OS Cryptomeria japonica.
FH Key Location/Qualifiers
FI peptide 1..21
FI protein 22..374
FI /label= signal_peptide
FI /label= mature_Cryj1
PI J06197768-A.
PI 19-JUN-1994.
PI 37-JAN-1993: 311116.
PI 07-JAN 1993: JP-001116.
PA (MEIC) MEIC: SEIKA KAISHA.
DR WPI: 94-288680/33.
DR N-PSDB: Q71601.
PI Sugi (Japanese cedar) pollen antigen Cryj1 - is useful for
PI diagnosis, treatment and prevention of sugi pollinosis.
PS Claim 2: Page 5-7: 9pp: Japanese.
CC The coding sequence for the Japanese cedar ("sugi") pollen allergen
CC Cryj1 was isolated from a cDNA library prepared from polyA mRNA. All
CC or part of the Cryj1 protein can be used for diagnosis, treatment
CC and prevention of sugi pollinosis.
SQ Sequence 374 AA:

Query Match 100.0%; Score 64; DB 1; Length 374;
Best Local Similarity 100.0%; Pred. No. 1,256-00;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DL 130 FIKRVSNVI 138
      11111111
QY : FIKRVSNVI 9

Search completed: Mon Jun 19 16:06:55 2000
Job time : 5 secs.

/note= "mat ... Cry j 1"
PI WO9301213-A.
PI 21-JAN-1993.
PI 10-JUL-1992: U05661.
PI 15-JUL-1991: US-729134.
PI 15-JUL-1991: US-730452.
PA (IMMU-) IMMUNOLOGIC PHARM CORP.
PI Bond JF., Griffith JG., Pollack J
PI WPI: 93-045434/05.
DR N-PSDB: Q35304.
PI Nucleic acid sequence encoding Cryptomeria japonica allergen
PI for the diagnosis treatment and prevention of allergic reactions
PI to Japanese cedar pollen.
PS Claim 11: Page 42: 69pp: English.
CC Fresh pollen and staminate
CC Cryptomeria japonica (Japanese cedar) tree RNA was prepd. and seq
CC to synthesize cDNA. The cDNA was subjected to successive rounds of
CC PCR to yield a full length
CC fragment of it may be used
CC for detecting, treating and preventing an
CC allergic response to Japanese cedar pollen allergen. It is capable of
CC modifying both the B and T cell response to Cry j 1 and T cell response
CC to a Cry j 1 antigen.
SQ Sequence 374 AA:

Query Match 100.0%; Score 64; DB 1; Length 374;
Best Local Similarity 100.0%; Pred. No. 1,256-00;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DL 130 FIKRVSNVI 138
      11111111
QY : FIKRVSNVI 9

Search completed: Mon Jun 19 16:06:55 2000
Job time : 5 secs.

```



CC TOPOLOGY: LINEAR  
CC MOLECULE TYPE: PROTEIN  
SQ SEQUENCE 317 AA: 37355 MW: 525334 CN:

Query Match 78.1% Score 50: DB 3: Length 317:  
Best Local Similarity 66.7% Pred. No. 4.45e+01:  
Matches 6: Conservative 2: Mismatches 1: Indels 0: Gaps 0:

DB 72 FPKRVDEVI 80

QY 1 FPKRVSNVI 9

RESULT 2  
ID US-08-951-889-2 STANDARD: PRT: 317 AA:

XX XXXXXX

Sequence 2, Application US/08951889

Sequence 2, Application US/08951889

Patent No. 6008332

GENERAL INFORMATION:

APPLICANT: Mathur, E. et al.

TITLE OF INVENTION: Carboxymethyl Cellulase from

THERMOTOGA MARITIMA

NUMBER OF SEQUENCES: 4

CORRESPONDENCE ADDRESS:

ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,

ADDRESSEE: CECCHI, STEWART & OLSTEIN

CITY: ROSELAND

STATE: NEW JERSEY

COUNTRY: USA

ZIP: 07068

COMPUTER READABLE FORM:

MEDIUM TYPE: 3.5 INCH DISKETTE

COMPUTER: IBM PS/2

OPERATING SYSTEM: MS-DOS

SOFTWARE: WORD PERFECT 5.1

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/951-889

FILING DATE: August 23, 1995

CLASSIFICATION: 4.35

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/951-889

FILING DATE: August 23, 1995

NAME: FERRARO, GREGORY D.

REGISTRATION NUMBER: 36,134

REFERENCE/DOCKET NUMBER: 331400-20

TELEPHONE: 201-994-1700

TELEFAX: 201-994-1744

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 317 AMINO ACIDS

TYPE: AMINO ACID

STRANDEDNESS:

TOPOLOGY: LINEAR

MOLECULE TYPE: PROTEIN

SEQUENCE 317 AA: 37355 MW: 525334 CN:

Query Match 78.1% Score 50: DB 3: Length 317:

Best Local Similarity 66.7% Pred. No. 4.45e+01:

Matches 6: Conservative 2: Mismatches 1: Indels 0: Gaps 0:

DB 72 FPKRVDEVI 80

QY 1 FPKRVSNVI 9

RESULT 3  
ID US-09-066-075-2 STANDARD: PRT: 317 AA:

XX XXXXXX

Sequence 2, Application US/09066075

Sequence 2, Application US/09066075

Patent No. 5925749

GENERAL INFORMATION:

APPLICANT: Mathur, E. et al.

TITLE OF INVENTION: Carboxymethyl Cellulase from *Thermotoga Maritima*

NUMBER OF SEQUENCES: 4

CORRESPONDENCE ADDRESS:

ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,

ADDRESSEE: CECCHI, STEWART & OLSTEIN

CITY: ROSELAND

STATE: NEW JERSEY

COUNTRY: USA

ZIP: 07068

COMPUTER READABLE FORM:

MEDIUM TYPE: 3.5 INCH DISKETTE

COMPUTER: IBM PS/2

OPERATING SYSTEM: MS-DOS

SOFTWARE: WORD PERFECT 5.1

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/066-075

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/518-615

FILING DATE: August 23, 1995

NAME: FERRARO, GREGORY D.

REGISTRATION NUMBER: 36,134

REFERENCE/DOCKET NUMBER: 331400-20

TELEPHONE: 201-994-1700

TELEFAX: 201-994-1744

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 317 AMINO ACIDS

TYPE: AMINO ACID

STRANDEDNESS:

TOPOLOGY: LINEAR

MOLECULE TYPE: PROTEIN

SEQUENCE 317 AA: 37355 MW: 525334 CN:

Query Match 78.1% Score 50: DB 2: Length 317:

Best Local Similarity 66.7% Pred. No. 4.45e+01:

Matches 6: Conservative 2: Mismatches 1: Indels 0: Gaps 0:

DB 72 FPKRVDEVI 80

QY 1 FPKRVSNVI 9

RESULT 4  
ID US-08-459-263-4 STANDARD: PRT: 255 AA:

XX XXXXXX

Sequence 4, Application US/08459263

Sequence 4, Application US/08459263

Patent No. 5792642

GENERAL INFORMATION:

CC APPLICANT: RASMUSSEN, Beth A  
CC APPLICANT: TALLY, Francis P  
CC TITLE OF INVENTION: DNA SEQUENCES AND AMINO ACID SEQUENCES  
CC TITLE OF INVENTION: OF CLASS B BETA-LACTAMASE ENZYMES FROM BACTERIOIDES  
CC TITLE OF INVENTION: FRAGILIS  
CC NUMBER OF SEQUENCES: 9  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Darby & Darby PC  
CC STREET: 805 Third Avenue  
CC CITY: New York  
CC STATE: NY  
CC COUNTRY: US  
CC ZIP: 10022  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: Patent Release #1.0, Version #1.30  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/459,263  
CC FILING DATE: 02-JUN-1995  
CC CLASSIFICATION: 435  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Robinson, Joseph R  
CC REGISTRATION NUMBER: 33,448  
CC REFERENCE/DOCKET NUMBER: 0646/1B026-US2  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: 212-527-7700  
CC TELEFAX: 212-753-6237  
CC TELEX: 236687  
CC INFORMATION FOR SEQ ID NO: 4:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 255 amino acids  
CC TYPE: amino acid  
CC STRANDEDNESS: single  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: peptide  
CC ORIGINAL SOURCE:  
CC ORGANISM: Bacillus cereus  
CC IMMEDIATE SOURCE:  
CC CLONE: B-LACTAMASE  
CC SEQUENCE 255 AA: 27847 MW: 349604 CN:  
  
Query Match 76.6% Score 49: DB 1: Length 255:  
Best Local Similarity 66.7% Pred. No. 5.76e-01:  
Matches 6: Conservative 2: Mismatches 1: Gaps 0:  
  
DB 103 FOKRVTDVI 111  
1111111111  
QY 1 FIKRVSNVI 9  
  
RESULT 5  
ID US-08-459-264-4 STANDARD: PRT: 255 AA:  
XX xxxxxx  
AC  
DT  
DE  
Sequence 4, Application US/08459264  
Patent No. 5705340  
GENERAL INFORMATION:  
CC APPLICANT: RASMUSSEN, Beth A  
CC APPLICANT: TALLY, Francis P  
CC APPLICANT: GLUZMAN, Yakov  
CC TITLE OF INVENTION: DNA SEQUENCES AND AMINO ACID SEQUENCES  
CC TITLE OF INVENTION: OF CLASS B BETA-LACTAMASE ENZYMES FROM BACTERIOIDES  
CC NUMBER OF SEQUENCES: 9  
CC CORRESPONDENCE ADDRESS:

CC ADDRESSEE: Darby & Darby PC  
CC STREET: 805 Third Avenue  
CC CITY: New York  
CC STATE: NY  
CC COUNTRY: US  
CC ZIP: 10022  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: Patent Release #1.0, Version #1.30  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/459,264  
CC FILING DATE: 02-JUN-95  
CC CLASSIFICATION: 435  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Robinson, Joseph R  
CC REGISTRATION NUMBER: 33,448  
CC REFERENCE/DOCKET NUMBER: 0646/1B026-US3  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: 212-527-7700  
CC TELEFAX: 212-753-6237  
CC TELEX: 236687  
CC INFORMATION FOR SEQ ID NO: 4:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 255 amino acids  
CC TYPE: amino acid  
CC STRANDEDNESS: single  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: peptide  
CC ORIGINAL SOURCE:  
CC ORGANISM: Bacillus cereus  
CC IMMEDIATE SOURCE:  
CC CLONE: B-LACTAMASE  
CC SEQUENCE 255 AA: 27847 MW: 349604 CN:  
  
Query Match 76.6% Score 49: DB 1: Length 255:  
Best Local Similarity 66.7% Pred. No. 5.76e-01:  
Matches 6: Conservative 2: Mismatches 1: Gaps 0:  
  
DB 103 FOKRVTDVI 111  
1111111111  
QY 1 FIKRVSNVI 9  
  
RESULT 6  
ID US-08-459-264-4 STANDARD: PRT: 424 AA:  
XX xxxxxx  
AC  
DT  
DE  
Sequence 1, Application US/08978182  
Patent No. 5849556  
GENERAL INFORMATION:  
CC APPLICANT: Hillman, Jennifer L.  
CC APPLICANT: Yue, Henry  
CC APPLICANT: Guegler, Karl J.  
CC APPLICANT: Kaser, Matthew  
CC APPLICANT: Mathur, Preete  
CC TITLE OF INVENTION: HUMAN GROWTH-RELATED CDC10 HOMOLOG  
CC NUMBER OF SEQUENCES: 5  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Incyte Pharmaceuticals, Inc.  
CC STREET: 3174 Porter Drive  
CC CITY: Palo Alto  
CC STATE: CA  
CC COUNTRY: USA  
CC ZIP: 94304  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Diskette

CC COMPUTER: IBM Compatible  
 CC OPERATING SYSTEM: DOS  
 CC SOFTWARE: FastSeq for Windows Version 2.0  
 CC CURRENT APPLICATION DATA:  
 CC APPLICATION NUMBER: US/08/978,182  
 CC FILING DATE: 08/978,182  
 CC REGISTRATION NUMBER: 435  
 CC CLASSIFICATION: 435  
 CC PRIOR APPLICATION DATA:  
 CC APPLICATION NUMBER:  
 CC FILING DATE:  
 CC ATTORNEY/AGENT INFORMATION:  
 CC NAME: Billings, Lucy J.  
 CC REGISTRATION NUMBER: 36,749  
 CC REFERENCE/DOCKET NUMBER: PF-0426 US  
 CC TELECOMMUNICATION INFORMATION:  
 CC TELEPHONE: 650-855-0555  
 CC TELEFAX: 650-845-4156  
 CC TELEX:  
 CC INFORMATION FOR SEQ ID NO: 1:  
 CC SEQUENCE CHARACTERISTICS:  
 CC LENGTH: 424 amino acids  
 CC TYPE: amino acid  
 CC STRANDEDNESS: single  
 CC TOPOLOGY: linear  
 CC IMMEDIATE SOURCE:  
 CC LIBRARY: TLYMNO106  
 CC CLONE: 3003826  
 CC SEQUENCE 424 AA: 47734 MW: 90,865.0N:  
 CC  
 CC Query Match 73.4% Score 471 DB 20 Length 424:  
 CC Best Local Similarity 55.6% Pred No: 9,576-01:  
 CC Matches 5: Conservative 4: Mismatches 0: Indels 0: Gaps 0:  
 CC  
 CC Db 265 EMKRSKV 273  
 CC 1111111  
 CC QY 1 FIKRVSNVI 9  
 CC  
 CC RESULT 7  
 CC ID US-09-205-487-1 STANDARD: PFI: 424 AA.  
 CC AC xxxxxx  
 CC DT  
 CC XX  
 CC Sequence 1 Application: 08/02/1968:  
 CC Sequence 1 Application: 08/02/1968:  
 CC Patent No: 545,234  
 CC GENERAL INFORMATION:  
 CC APPLICANT: Billman, Jennifer D.  
 CC APPLICANT: Y. P. Henry  
 CC APPLICANT: Giesler, Karl J.  
 CC APPLICANT: Kaser, Matthew  
 CC APPLICANT: Mathur, Brooke  
 CC TITLE OF INVENTION: HUMAN GROWTH-RELATED CD-10 PROTEIN  
 CC NUMBER OF SEQUENCES: 1  
 CC CORRESPONDENCE ADDRESS:  
 CC ADDRESSEE: Inocyte Pharmaceuticals Inc.  
 CC STREET: 9174 Porter Drive  
 CC CITY: Palo Alto  
 CC STATE: CA  
 CC COUNTRY: USA  
 CC ZIP: 94304  
 CC COMPUTER READABLE FORM:  
 CC MEDIUM TYPE: Diskette  
 CC COMPUTER: IBM Compatible  
 CC OPERATING SYSTEM: DOS  
 CC SOFTWARE: FastSeq for Windows Version 2.0  
 CC CURRENT APPLICATION DATA:  
 CC APPLICATION NUMBER: US/09/205,481  
 CC FILING DATE:  
 CC CLASSIFICATION:

CC PRIOR APPLICATION DATA:  
 CC APPLICATION NUMBER: 08/978,182  
 CC FILING DATE:  
 CC ATTORNEY/AGENT INFORMATION:  
 CC NAME: Billings, Lucy J.  
 CC REGISTRATION NUMBER: 435  
 CC REFERENCE/DOCKET NUMBER: PF-0426 US  
 CC TELECOMMUNICATION INFORMATION:  
 CC TELEPHONE: 650-855-0555  
 CC TELEFAX: 650-845-4156  
 CC TELEX:  
 CC INFORMATION FOR SEQ ID NO: 1:  
 CC SEQUENCE CHARACTERISTICS:  
 CC LENGTH: 424 amino acids  
 CC TYPE: amino acid  
 CC STRANDEDNESS: single  
 CC TOPOLOGY: linear  
 CC IMMEDIATE SOURCE:  
 CC LIBRARY: TLYMNO106  
 CC CLONE: 3003826  
 CC SEQUENCE 424 AA: 47734 MW: 90,865.0N:  
 CC  
 CC Query Match 73.4% Score 471 DB 20 Length 424:  
 CC Best Local Similarity 55.6% Pred No: 9,576-01:  
 CC Matches 5: Conservative 4: Mismatches 0: Indels 0: Gaps 0:  
 CC  
 CC Db 265 EMKRSKV 273  
 CC 1111111  
 CC QY 1 FIKRVSNVI 9  
 CC  
 CC RESULT 8  
 CC ID FCI-US95-10398-80 STANDARD: PFI: 192 AA.  
 CC AC xxxxxx  
 CC DT  
 CC XX  
 CC Sequence 80 Application: 09/05/1998:  
 CC Sequence 80 Application: 09/05/1998:  
 CC GENERAL INFORMATION:  
 CC APPLICANT: BURKH, J. M. & P. H. AND  
 CC APPLICANT: PURCELL, R. J.  
 CC TITLE OF INVENTION: NOVEL HIF AND DETECTED  
 CC TITLE OF INVENTION: AMIN ACID SEQUENCES OF THE ENVELOPE 1 AG  
 CC TITLE OF INVENTION: NOVEL HIF AND DETECTED  
 CC TITLE OF INVENTION: AMIN ACID SEQUENCES OF THE ENVELOPE 1 AG  
 CC TITLE OF INVENTION: AMIN ACID SEQUENCES OF THE ENVELOPE 1 AG  
 CC TITLE OF INVENTION: AMIN ACID SEQUENCES OF THE ENVELOPE 1 AG  
 CC NUMBER OF SEQUENCES: 1  
 CC CORRESPONDENCE ADDRESS:  
 CC ADDRESSEE: MORGAN & FINNIGAN  
 CC STREET: 345 PARK AVENUE  
 CC CITY: NEW YORK  
 CC STATE: NEW YORK  
 CC COUNTRY: USA  
 CC ZIP: 10154  
 CC COMPUTER READABLE FORM:  
 CC MEDIUM TYPE: FLOPPY DISK  
 CC COMPUTER: IBM PC COMPATIBLE  
 CC OPERATING SYSTEM: PC-DOS/MS-DOS  
 CC SOFTWARE: WORDPERFECT 5.1  
 CC CURRENT APPLICATION DATA:  
 CC APPLICATION NUMBER: 09/085,428  
 CC FILING DATE: 15-AUG-1995  
 CC CLASSIFICATION:  
 CC PRIOR APPLICATION DATA:  
 CC APPLICATION NUMBER: 08/085,428  
 CC FILING DATE: 29 JUNE 1993  
 CC PRIOR APPLICATION DATA:  
 CC APPLICATION NUMBER: 08/290/655  
 CC FILING DATE: 15 AUGUST 1994  
 CC ATTORNEY/AGENT INFORMATION:



CC NAME: RICHARD W. BORK  
CC REGISTRATION NUMBER: 36,459  
CC REFERENCE/DOCKET NUMBER: 2026-4116  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (212) 758-4800  
CC TELEFAX: (212) 751-6849  
CC TELEX: 421792  
CC INFORMATION FOR SEQ ID NO: 80:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 192 amino acids  
CC TYPE: amino acid  
CC STRANDEDNESS: unknown  
CC TOPOLOGY: unknown  
CC ORIGINAL SOURCE:  
CC ORGANISM: homosapiens  
CC INDIVIDUAL ISOLATE: US10  
CC SEQUENCE 192 AA: 20855 MW: 200760 CN:

Query Match 70.3% Score 45: DB 4: Length 192:  
Best Local Similarity 44.4% Pred. No. 1: 58e-02:  
Matches 4: Conservative 5: Mismatches 0: Indels 0: Gaps 0:

Db 145 YVMRVPVEVI 153  
QY 1 FIKRVSNI 9

RESULT 9  
ID US-08-290-665A-80 STANDARD: PRT: 192 AA.

XX AC xxxxxx

XX DE

Sequence 80, Application US/08290665A

Sequence 80, Application US/08290665A

Patent No. 582852

GENERAL INFORMATION:

APPLICANT: BURKH, J., MILLER, R. H. AND

APPLICANT: PURCELL, R. H.

TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED

TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND

TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS

TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE

TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES

NUMBER OF SEQUENCES: 263

CORRESPONDENCE ADDRESS:

ADDRESSEE: MORGAN & FINNEGAN

STREET: 345 PARK AVENUE

CITY: NEW YORK

STATE: NEW YORK

COUNTRY: USA

ZIP: 10154

COMPUTER READABLE FORM:

MEDIUM TYPE: FLOPPY DISK

COMPUTER: IBM PC COMPATIBLE

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: WORDPERFECT 5.1

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/290,665A

FILING DATE: 15-AUG-1994

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: RICHARD W. BORK

REGISTRATION NUMBER: 36,459

REFERENCE/DOCKET NUMBER: 2026-4116

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 758-4800

TELEFAX: (212) 751-6849

TELEX: 421792

INFORMATION FOR SEQ ID NO: 80:

SEQUENCE CHARACTERISTICS:

CC LENGTH: 192 amino acids  
CC TYPE: amino acid  
CC STRANDEDNESS: unknown  
CC TOPOLOGY: unknown  
CC ORIGINAL SOURCE:  
CC ORGANISM: homosapiens  
CC INDIVIDUAL ISOLATE: US10  
CC SEQUENCE 192 AA: 20855 MW: 200760 CN:

Query Match 70.3% Score 45: DB 2: Length 192:  
Best Local Similarity 44.4% Pred. No. 1: 58e-02:  
Matches 4: Conservative 5: Mismatches 0: Indels 0: Gaps 0:

Db 145 YVMRVPVEVI 153  
QY 1 FIKRVSNI 9

RESULT 10  
ID US-08-086-428B-80 STANDARD: PRT: 192 AA.

XX AC xxxxxx

XX DE

XX DE

Sequence 80, Application US/08086428B

Sequence 80, Application US/08086428B

Patent No. 5514539

GENERAL INFORMATION:

APPLICANT: BURKH, J., MILLER, R. H. AND

APPLICANT: PURCELL, R. H.

TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED

TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 GENE

TITLE OF INVENTION: OF 51 ISOLATES OF HEPATITIS C AND THE USE

TITLE OF INVENTION: OF REAGENTS DERIVED FROM THESE SEQUENCES IN

TITLE OF INVENTION: DIAGNOSTIC METHODS AND VACCINES

NUMBER OF SEQUENCES: 159

CORRESPONDENCE ADDRESS:

ADDRESSEE: MORGAN & FINNEGAN

STREET: 345 PARK AVENUE

CITY: NEW YORK

STATE: NEW YORK

COUNTRY: USA

ZIP: 10154

COMPUTER READABLE FORM:

MEDIUM TYPE: FLOPPY DISK

COMPUTER: IBM PC COMPATIBLE

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: WORDPERFECT 5.1

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/086,428B

FILING DATE: 29-JUN-1993

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: RICHARD W. BORK

REGISTRATION NUMBER: 36,459

REFERENCE/DOCKET NUMBER: 2026-4070

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 758-4800

TELEFAX: (212) 751-6849

TELEX: 421792

INFORMATION FOR SEQ ID NO: 80:

SEQUENCE CHARACTERISTICS:

LENGTH: 192 amino acids

TYPE: amino acid

STRANDEDNESS: unknown

TOPOLOGY: unknown

ORIGINAL SOURCE:

ORGANISM: homosapiens

INDIVIDUAL ISOLATE: US10

SEQUENCE 192 AA: 20855 MW: 200760 CN:

Query Match 70.3% Score 45: DB 1: Length 192:  
 Best Local Similarity 44.4% Pred. No. 1.58e+02:  
 Matches 4: Conservative 5: Mismatches 0: Indels 0: Gaps 0:

DB 145 YVMKRPV1 153  
 QY 1 FIKRVSNV1 9

RESULT 1:  
 ID US-08-458-570-80 STANDARD: PRT: 192 AA  
 AC xxxxxx  
 DE Sequence 80, Application US/08458570  
 XX Sequence 80, Application US/08458570  
 CC Patent No. 5871962  
 CC GENERAL INFORMATION:  
 CC APPLICANT: RUKH, J., MILLER, R.H. AND  
 CC APPLICANT: PORCELL, R.H.  
 CC TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED  
 CC TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 GENE  
 CC TITLE OF INVENTION: OF 51 ISOLATES OF HEPATITIS C AND THE USE  
 CC TITLE OF INVENTION: OF REAGENTS DERIVED FROM THESE SEQUENCES IN  
 CC TITLE OF INVENTION: DIAGNOSTIC METHODS AND VACCINES  
 CC NUMBER OF SEQUENCES: 159  
 CC CORRESPONDENCE ADDRESS:  
 CC ADDRESSEE: MORGAN & FINNEGAN  
 CC STREET: 345 PARK AVENUE  
 CC CITY: NEW YORK  
 CC STATE: NEW YORK  
 CC COUNTRY: USA  
 CC ZIP: 10154  
 CC COMPUTER READABLE FORM:  
 CC MEDIUM TYPE: FLOPPY DISK  
 CC COMPUTER: IBM PC COMPATIBLE  
 CC OPERATING SYSTEM: PC-DOS/MS-DOS  
 CC SOFTWARE: WORDPERFECT 5.1  
 CC CURRENT APPLICATION DATA:  
 CC APPLICATION NUMBER: US/08/458,570  
 CC FILING DATE: JUN-1995  
 CC CLASSIFICATION: 424  
 CC PRIOR APPLICATION DATA:  
 CC APPLICATION NUMBER: 08/186,428  
 CC FILING DATE: JUN-1993  
 CC CLASSIFICATION: 424  
 CC ATTORNEY/AGENT INFORMATION:  
 CC NAME: RICHARD W. BORK  
 CC REGISTRATION NUMBER: 36,459  
 CC REFERENCE/DOCKET NUMBER: 2026-4070US:  
 CC TELECOMMUNICATION INFORMATION:  
 CC TELEPHONE: (212) 758-4800  
 CC TELEFAX: (212) 751-6849  
 CC TELEX: 421792  
 CC INFORMATION FOR SEQ ID NO: 80:  
 CC SEQUENCE CHARACTERISTICS:  
 CC LENGTH: 192 amino acids  
 CC TYPE: amino acid  
 CC STRANDEDNESS: unknown  
 CC TOPOLOGY: unknown  
 CC ORIGINAL SOURCE:  
 CC ORGANISM: homoplans  
 CC INDIVIDUAL ISOLATE: US10  
 CC SEQUENCE 192 AA: 20855 MW: 200760 CN:

Query Match 70.3% Score 45: DB 2: Length 192:  
 Best Local Similarity 44.4% Pred. No. 1.58e+02:  
 Matches 4: Conservative 5: Mismatches 0: Indels 0: Gaps 0:

DB 145 YVMKRPV1 153

QY 1 FIKRVSNV1 9

RESULT 12  
 ID US-09-393-985-18 STANDARD: PRT: 206 AA  
 AC xxxxxx  
 DE Sequence 18, Application US/08393985  
 XX Sequence 18, Application US/08393985  
 CC Patent No. 5693676  
 CC GENERAL INFORMATION:  
 CC APPLICANT: Scheller, Richard H.  
 CC TITLE OF INVENTION: Methods and Compositions for Modulation  
 CC TITLE OF INVENTION: of Vesicular Release  
 CC NUMBER OF SEQUENCES: 35  
 CC CORRESPONDENCE ADDRESS:  
 CC ADDRESSEE: Dehlinger & Associates  
 CC STREET: 350 Cambridge Avenue, Suite 250  
 CC CITY: Palo Alto  
 CC STATE: CA  
 CC COUNTRY: USA  
 CC ZIP: 94306  
 CC COMPUTER READABLE FORM:  
 CC MEDIUM TYPE: Floppy disk  
 CC COMPUTER: IBM PC compatible  
 CC OPERATING SYSTEM: PC-DOS/MS-DOS  
 CC SOFTWARE: PatentIn Release 1.0, Version 1.25  
 CC CURRENT APPLICATION DATA:  
 CC APPLICATION NUMBER: US/09/393,985  
 CC FILING DATE:  
 CC CLASSIFICATION: 435  
 CC ATTORNEY/AGENT INFORMATION:  
 CC NAME: Sholtz, Charles K.  
 CC REGISTRATION NUMBER: 43,416  
 CC REFERENCE/DOCKET NUMBER: 8600-0152  
 CC TELECOMMUNICATION INFORMATION:  
 CC TELEPHONE: (415) 324-9880  
 CC TELEFAX: (415) 324-0966  
 CC INFORMATION FOR SEQ ID NO: 19:  
 CC SEQUENCE CHARACTERISTICS:  
 CC LENGTH: 206 amino acids  
 CC TYPE: amino acid  
 CC TOPOLOGY: linear  
 CC MOLECULE TYPE: protein  
 CC SEQUENCE 206 AA: 23915 MW: 177393 CN:

Query Match 70.3% Score 45: DB 1: Length: 206:  
 Best Local Similarity 71.4% Pred. No. 1.58e+02:  
 Matches 5: Conservative 2: Mismatches 0: Indels 0: Gaps 0:

DB 133 FIKRVIN 139  
 QY 1 FIKRVSN 7

RESULT 13  
 ID US-08-118-270-76 STANDARD: PRT: 296 AA  
 AC xxxxxx  
 DE Sequence 76, Application US/08118270  
 XX Sequence 76, Application US/08118270  
 CC Patent No. 5508384  
 CC GENERAL INFORMATION:  
 CC APPLICANT: Murphy, Randall B.

```

CC APPLICANT: Schuster, David I.
CC TITLE OF INVENTION: POLYPEPTIDES OF G-COUPLED PROTEIN
CC TITLE OF INVENTION: RECEPTORS, AND COMPOSITIONS AND METHODS THEREOF
CC NUMBER OF SEQUENCES: 348
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: BROWDY AND NEIMARK
CC STREET: 419 Seventh Street, N.W., Suite 300
CC CITY: Washington
CC STATE: D.C.
CC COUNTRY: USA
CC ZIP: 20004
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patent Release #1.0. Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/118,270
CC FILING DATE: 09-SEP-1993
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 07/943,236
CC FILING DATE: 10-SEP-1992
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Townsend, Kevin G.
CC REGISTRATION NUMBER: 34,033
CC REFERENCE/DOCKET NUMBER: MURPHY-2A
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 202-628-5197
CC TELEFAX: 202-737-3528
CC TELEX: 248633
CC INFORMATION FOR SEQ ID NO: 76:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 298 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC SEQUENCE: 298 AA; 33548 MW; 467536 CN;

Query Match: 70.3%; Score 45; DB 1; Length 298;
Best Local Similarity: 33.3%; Pred. No.: 55e+02;
Matches: 3; Conservative: 5; Mismatches: 1; Indels: 0; Gaps: 0;

ID# 7: YVRRVSRIIV 79
QY ::::II::
1 FIREVSNVI 9

RESULT 14
ID# PC-TUS93-08528-76 STANDARD: PKT: 298 AA.
XX xxxxxx
XX AC
XX XX
D:
XX
DE Sequence 76, Application PC/TUS9308528
XX
CC Sequence 76, Application PC/TUS9308528
CC GENERAL INFORMATION:
CC APPLICANT: New York University
CC TITLE OF INVENTION: POLYPEPTIDES OF G-COUPLED PROTEIN
CC TITLE OF INVENTION: RECEPTORS, AND COMPOSITIONS AND METHODS THEREOF
CC NUMBER OF SEQUENCES: 348
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: BROWDY AND NEIMARK
CC STREET: 419 Seventh Street, N.W., Suite 300
CC CITY: Washington
CC STATE: D.C.
CC COUNTRY: USA
CC ZIP: 20004
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible

```

```
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.3, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: PCT/US93/08528
CC FILING DATE: 09-SEP-1993
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 07/943,235
CC FILING DATE: 10-SEP-1992
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Townsend, Kevin G.
CC REGISTRATION NUMBER: 34,033
CC REFERENCE/DOCKET NUMBER: MURPHY-2 PCT
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 202-628-5197
CC TELEFAX: 202-737-3528
CC TELEX: 248633
CC INFORMATION FOR SEQ ID NO: 76:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 298 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC SEQUENCE 298 AA: 3354B MS: 45753E CN:
CC
CC Query Match: 70.3% Score 45; DB 4; Length 298;
CC Best Local Similarity 33.3%; Pred. No. 1.58e+02;
CC Matches 3; Conservative 5; Mismatches 1; Indels 0; Gaps 0;
CC
Db 71 YVHRVSRIIV 79
QY ::::
X XXXXXX
AC
XX
XX
DT
XX
XX
DE Sequence 9, Application US/06,46762A
XX
CC Sequence 9, Application US/08946762A
CC Patent No. 5994072
CC GENERAL INFORMATION:
CC APPLICANT: Lem, Joseph S
CC APPLICANT: Burrows, Lori
CC APPLICANT: Chaffetz, Deborah
CC APPLICANT: de Kleijn, Teresa
CC TITLE OF INVENTION: NO. 5994072el Proteins Involved in the Synthesis and Assem
CC TITLE OF INVENTION: Of O-Antigen in Pseudomonas Aeruginosa
CC FILE REFERENCE: 6580-089
CC CURRENT APPLICATION NUMBER: US/08/846,762A
CC CURRENT FILING DATE: 1997-04-30
CC NUMBER OF SEQ ID NOS: 100
CC SOFTWARE: PatentIn Ver. 2.0
CC SEQ ID NO 9
CC LENGTH: 316
CC TYPE: PRT
CC ORGANISM: Pseudomonas aeruginosa
CC SEQUENCE 316 AA: 33805 MW: 521677 CN:
CC
CC Query Match: 70.3% Score 45; DB 2; Length 316;
CC Best Local Similarity 55.0%; Pred. No. 1.58e+02;
CC Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
CC
Db 5 FIRRVARVL 13
QY ::::
X FIRRVSNVI 9
```

Search completed: Tue Jun 20 13:35:15 2000

Jot time : 5 secs

---





```

CC      TOPOLOGY: linear
CC      MOLECULE TYPE: peptide
CC      FRAGMENT TYPE: internal
SQ      SEQUENCE 20 AA: 2080 MW: 2304 CN:

Query Match: 100.0% Score 64: DB 4: Length 20:
Best Local Similarity 100.0% Pred. No. 2.93e+00:
Matches 9: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

Db 9 FIKRVSNVI 17
QY 1 FIKRVSNVI 9

RESULT 4
ID US-08-226-248A-36 STANDARD: PRT: 20 AA:
XX xxxxxx
AC
DE
DT
XX
XX
XX
XX
XX
DE
CC Sequence 36, Application US/08226248A
CC Sequence 36, Application US/08226248A
CC GENERAL INFORMATION:
CC APPLICANT: Griffith, Irwin J.;
CC APPLICANT: Pollock, Joanne;
CC APPLICANT: Bond, Julian F.;
CC APPLICANT: Garman, Richard D.;
CC APPLICANT: Kuo, Mei-Chang;
CC APPLICANT: Brauer, Andrew;
CC APPLICANT: Exley, Mark A.;
CC APPLICANT: Powers, Steven P.;
CC APPLICANT: Yeung, Siu-mei H.;
CC APPLICANT: Kuo, Mei-Chang;
CC APPLICANT: Brauer, Andrew;
CC APPLICANT: Exley, Mark A.;
CC APPLICANT: Powers, Steven P.;
CC TITLE OF INVENTION: Allergenic Proteins And Peptides From
CC TITLE OF INVENTION: Japanese Cedar Pollen
CC NUMBER OF SEQUENCES: 201
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Immunologic Pharmaceutical Corporation, Inc.
CC STREET: 610 Lincoln St
CC CITY: Waltham
CC STATE: MA
CC COUNTRY: USA
CC ZIP: 02154
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patent Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/226,248A
CC FILING DATE: April 8, 1994
CC CLASSIFICATION: 424
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 07/938,990
CC FILING DATE: September 1, 1992
CC APPLICATION NUMBER: PCT/US93/00139
CC FILING DATE: January 15, 1993
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Darlene A. Vansone
CC REGISTRATION NUMBER: 35,729
CC REFERENCE/DOCKET NUMBER: 025.5 US (IMI-C28CP)
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (617) 466-6000
CC TELEFAX: (617) 466-6040
CC INFORMATION FOR SEQ ID NO: 36:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 20 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC FRAGMENT TYPE: internal
SQ SEQUENCE 20 AA: 2080 MW: 2304 CN:

Query Match: 100.0% Score 64: DB 10: Length 20:
Best Local Similarity 100.0% Pred. No. 2.93e+00:
Matches 9: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

Db 9 FIKRVSNVI 17
QY 1 FIKRVSNVI 9

Query Match: 100.0% Score 64: DB 7: Length 20:
Best Local Similarity 100.0% Pred. No. 2.93e+00:
Matches 9: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

Db 9 FIKRVSNVI 17
QY 1 FIKRVSNVI 9

```

QY 1 FIKRVSNVI 9

RESULT 4  
ID US-08-350-225-36 STANDARD: PPT: 20 AA.

XX AC xxxxxx

XX DI

XX XX

DE Sequence 36, Application US/08350225

XX Sequence 36, Application US/08350225

CC GENERAL INFORMATION:

CC APPLICANT: Griffith, Irwin J.

CC APPLICANT: Pollock, Joanne,

CC APPLICANT: Bond, Julian F.

CC APPLICANT: Garman, Richard D.

CC APPLICANT: Kuo, Mei-Chang.

CC APPLICANT: Yeung, Siu-mei H.

CC APPLICANT: Brauer, Andrew.

CC APPLICANT: Exley, Mark A.

CC APPLICANT: Powers, Steven P.

CC APPLICANT: Griffith, Irwin J.

CC TITLE OF INVENTION: Allergenic Proteins And Peptides From

CC TITLE OF INVENTION: Japanese Cedar Pollen

CC NUMBER OF SEQUENCES: 261

CC CORRESPONDENCE ADDRESS:

CC ADDRESSEE: Immunologic Pharmaceutical Corporation, Inc.

CC STREET: 610 Lincoln St

CC CITY: Waltham

CC STATE: MA

CC COUNTRY: USA

CC ZIP: 02154

CC COMPUTER READABLE FORM:

CC MEDIUM TYPE: Floppy disk

CC COMPUTER: IBM PC compatible

CC OPERATING SYSTEM: PC-DOS/MS-DOS

CC SOFTWARE: Patent Release #1.0, Version #1.25

CC CURRENT APPLICATION DATA:

CC APPLICATION NUMBER: US/08/350.225

CC FILING DATE: December 6, 1994

CC CLASSIFICATION: 424

CC PRIOR APPLICATION DATA:

CC APPLICATION NUMBER: 08/226,248

CC FILING DATE: April 8, 1994

CC APPLICATION NUMBER: 07/538,994

CC FILING DATE: September 1, 1992

CC APPLICATION NUMBER: 08/0593/0139

CC FILING DATE: January 15, 1995

CC ATTORNEY/AGENT INFORMATION:

CC NAME: Darlene A. Vastore

CC REGISTRATION NUMBER: 35,729

CC REFERENCE/DOCKET NUMBER: 025 6 US (IM1-028CP2)

CC TELECOMMUNICATION INFORMATION:

CC TELEPHONE: (617) 466-6000

CC TELEFAX: (617) 466-6040

CC INFORMATION FOR SEQ ID NO: 36:

CC SEQUENCE CHARACTERISTICS:

CC LENGTH: 20 amino acids

CC TYPE: amino acid

CC TOPOLOGY: linear

CC MOLECULE TYPE: peptide

CC FRAGMENT TYPE: internal

CC SEQUENCE 20 AA: 2080 MW: 2304 CN:

Query Match 100.0% Score 64: DB 8: Length 20:

Best Local Similarity 100.0% Pred. No. 2 93e-00:

Matches 9: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

DB 9 FIKRVSNVI 17

XX 11111111

QY 1 FIKRVSNVI 9

RESULT 7

ID US-09-467-697-36 STANDARD: PPT: 20 AA.

XX AC xxxxxx

XX DI

XX XX

DE Sequence 36, Application US/08467697

XX Sequence 36, Application US/08467697

CC GENERAL INFORMATION:

CC APPLICANT: Griffith, Irwin J.

CC APPLICANT: Pollock, Joanne,

CC APPLICANT: Bond, Julian F.

CC APPLICANT: Garman, Richard D.

CC APPLICANT: Kuo, Mei-Chang.

CC APPLICANT: Yeung, Siu-mei H.

CC APPLICANT: Brauer, Andrew.

CC APPLICANT: Exley, Mark A.

CC APPLICANT: Powers, Steven P.

CC APPLICANT: Griffith, Irwin J.

CC TITLE OF INVENTION: Allergenic Proteins And Peptides From

CC TITLE OF INVENTION: Japanese Cedar Pollen

CC NUMBER OF SEQUENCES: 261

CC CORRESPONDENCE ADDRESS:

CC ADDRESSEE: Immunologic Pharmaceutical Corporation, Inc.

CC STREET: 610 Lincoln St

CC CITY: Waltham

CC STATE: MA

CC COUNTRY: USA

CC ZIP: 02154

CC COMPUTER READABLE FORM:

CC MEDIUM TYPE: Floppy disk

CC COMPUTER: IBM PC compatible

CC OPERATING SYSTEM: PC-D /MS-DOS

CC SOFTWARE: Patent Release #1.0, Version #1.25

CC CURRENT APPLICATION DATA:

CC APPLICATION NUMBER: US/09/467.697

CC FILING DATE: June 6, 1995

CC CLASSIFICATION: 424

CC PRIOR APPLICATION DATA:

CC APPLICATION NUMBER: 08/350.225

CC FILING DATE: December 6, 1994

CC ATTORNEY/AGENT INFORMATION:

CC NAME: Darlene A. Vastore

CC REGISTRATION NUMBER: 35,729

CC REFERENCE/DOCKET NUMBER: 025 6 US (IM1-028CP2)

CC TELECOMMUNICATION INFORMATION:

CC TELEPHONE: (617) 227-7400

CC TELEFAX: (617) 227-7441

CC INFORMATION FOR SEQ ID NO: 36:

CC SEQUENCE CHARACTERISTICS:

CC LENGTH: 20 amino acids

CC TYPE: amino acid

CC TOPOLOGY: linear

CC MOLECULE TYPE: peptide

CC FRAGMENT TYPE: internal

CC SEQUENCE 20 AA: 2080 MW: 2304 CN:

Query Match 100.0% Score 64: DB 10: Length 20:

Best Local Similarity 100.0% Pred. No. 2 93e-00:

Matches 9: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

DB 9 FIKRVSNVI 17

XX 11111111

QY 1 FIKRVSNVI 9

RESULT 8

ID US-09-142-524-4 STANDARD: PPT: 31 AA.

XX AC xxxxxx



XX DE Query Match 100.0% Score 64: DB 17: Length 31:  
XX DE Best Local Similarity 100.0%: Pred. No. 2.93e-00:  
XX DE Matches 9: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

XX DE Sequence 4, Application US/09142524  
CC GENERAL INFORMATION:  
CC APPLICANT: Sone, Toshio  
CC APPLICANT: Kume, Akinori  
CC APPLICANT: Kairiki, Kazuo  
CC APPLICANT: Iwama, Akiko  
CC APPLICANT: Kino, Kohsuke  
CC TITLE OF INVENTION: Peptide-based Immunotherapeutic Agent For Treating  
CC TITLE OF INVENTION: Allergic Diseases  
CC FILE REFERENCE: Docket No. SPO-103  
CC CURRENT FILING DATE: 1999-01-04  
CC EARLIER APPLICATION NUMBER: 8/80/702  
CC EARLIER FILING DATE: 1996-03-10  
CC EARLIER APPLICATION NUMBER: PCT/JP97/00740  
CC EARLIER FILING DATE: 1997-03-10  
CC NUMBER OF SEQ ID NOS: 5  
CC SOFTWARE: PatentIn Ver. 2.0  
CC SEQ ID NO 4  
CC LENGTH: 31  
CC TYPE: PRT  
CC ORGANISM: Unknown  
CC FEATURE:  
CC OTHER INFORMATION: Description of Unknown Organism:peptide  
SQ SEQUENCE 31 AA: 3561 MW: 4960 CN:

DB 23 FIKRVSNVI 31  
QY 1 FIKRVSNVI 9

RESULT 9  
XX DE US-09-142-524-5 STANDARD: PRI: 31 AA:  
XX AC xxxxxx  
XX PT  
XX

DE Sequence 5, Application US/09142524  
XX DE Query Match 100.0% Score 64: DB 17: Length 31:  
XX DE Best Local Similarity 100.0%: Pred. No. 2.93e-00:  
XX DE Matches 9: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

XX DE Sequence 4, Application US/09142524  
CC GENERAL INFORMATION:  
CC APPLICANT: Sone, Toshio  
CC APPLICANT: Kume, Akinori  
CC APPLICANT: Kairiki, Kazuo  
CC APPLICANT: Iwama, Akiko  
CC APPLICANT: Kino, Kohsuke  
CC TITLE OF INVENTION: Peptide-based Immunotherapeutic Agent For Treating  
CC TITLE OF INVENTION: Allergic Diseases  
CC FILE REFERENCE: Docket No. SPO-103  
CC CURRENT FILING DATE: 1999-01-04  
CC EARLIER APPLICATION NUMBER: 8/80/702  
CC EARLIER FILING DATE: 1996-03-10  
CC EARLIER APPLICATION NUMBER: PCT/JP97/00740  
CC EARLIER FILING DATE: 1997-03-10  
CC NUMBER OF SEQ ID NOS: 5  
CC SOFTWARE: PatentIn Ver. 2.0  
CC SEQ ID NO 5  
CC LENGTH: 31  
CC TYPE: PRT  
CC ORGANISM: Unknown  
CC FEATURE:  
CC OTHER INFORMATION: Description of Unknown Organism:peptide

SQ SEQUENCE 31 AA: 3585 MW: 4902 CN:

Query Match 100.0% Score 64: DB 17: Length 31:  
Best Local Similarity 100.0%: Pred. No. 2.93e-00:  
Matches 9: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

DB 23 FIKRVSNVI 31  
QY 1 FIKRVSNVI 9

RESULT 10  
XX DE US-07-938-990A-62 STANDARD: PRI: 60 AA:  
XX AC xxxxxx  
XX DT  
XX DE  
XX DE Sequence 62, Application US/07938990A  
XX DE Sequence 62, Application US/07938990A  
XX DE GENERAL INFORMATION:  
XX DE APPLICANT: Griffith, Irwin J.  
XX DE APPLICANT: Pollock, Joanne  
XX DE APPLICANT: Bond Julian  
XX DE TITLE OF INVENTION: Allergenic Proteins And Peptides From  
XX DE TITLE OF INVENTION: Japanese Cedar Pollen  
XX DE NUMBER OF SEQUENCES: 70  
XX DE CORRESPONDENCE ADDRESS:  
XX DE ADDRESSEE: Lahive & Cockfield  
XX DE STREET: Sixty State Street  
XX DE CITY: Boston  
XX DE STATE: MA  
XX DE COUNTRY: USA  
XX DE ZIP: 02109  
XX DE COMPUTER READABLE FORM  
XX DE MEDIUM TYPE: F.cop disk  
XX DE COMPUTER: IBM PC compatible  
XX DE OPERATING SYSTEM: PC-DOS/MS-DOS  
XX DE SOFTWARE: PatentIn Release #1.0, Version #1.25  
XX DE CURRENT APPLICATION DATA:  
XX DE APPLICATION NUMBER: US/07938,990A  
XX DE FILING DATE: 19920901  
XX DE CLASSIFICATION: 435  
XX DE PRIOR APPLICATION DATA:  
XX DE APPLICATION NUMBER: 07/741,452  
XX DE FILING DATE: July 15, 1991  
XX DE PRIOR APPLICATION DATA:  
XX DE APPLICATION NUMBER: 07/741,452  
XX DE FILING DATE: July 16, 1991  
XX DE ATTORNEY/AGENT INFORMATION:  
XX DE NAME: Amy E. Mandragoras  
XX DE REGISTRATION NUMBER: 35,207  
XX DE REFERENCE/DOCKET NUMBER: IPC-025CC (IM-028)  
XX DE TELEPHONE: (617) 227-7400  
XX DE TELEFAX: (617) 227-5941  
XX DE INFORMATION FOR SEQ ID NO: 62:  
XX DE SEQUENCE CHARACTERISTICS:  
XX DE LENGTH: 60 amino acids  
XX DE TYPE: AMINO ACID  
XX DE TOPOLOGY: linear  
XX DE MOLECULE TYPE: peptide  
XX DE FRAGMENT TYPE: internal  
SQ SEQUENCE 60 AA: 6644 MW: 19464 CN:

Query Match 100.0% Score 64: DB 4: Length 60:  
Best Local Similarity 100.0%: Pred. No. 2.93e-00:  
Matches 9: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

DB 49 FIKRVSNVI 57  
QY 1 FIKRVSNVI 9

```

RESULT 11
ID US-08-468-940-62 STANDARD: PRT: 60 AA.
XX
AC xxxxxx
XX
DT
XX
DE Sequence 52, Application: US/08468940
XX
CC Sequence 52, Application US/08468940
CC GENERAL INFORMATION:
CC APPLICANT: Griffith, Irwin J.:
CC APPLICANT: Pollock, Joanne:
CC APPLICANT: Bond, Julian F.:
CC APPLICANT: Garman, Richard D.:
CC APPLICANT: Kuo, Mei-Chang:
CC APPLICANT: Yeung, Siu-mei H.:
CC APPLICANT: Brauer, Andrew:
CC APPLICANT: Exley, Mark A.:
CC APPLICANT: Powers, Steven P.:
CC TITLE OF INVENTION: Allergenic Proteins And Peptides From
CC TITLE OF INVENTION: Japanese Cedar Pollen
CC NUMBER OF SEQUENCES: 251
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Immunologic Pharmaceutical Corporation, Inc.
CC STREET: 610 Lincoln St
CC CITY: Waltham
CC STATE: MA
CC COUNTRY: USA
CC ZIP: 02154
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patent Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08468-940
CC FILING DATE:
CC CLASSIFICATION: 424
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/350,225
CC FILING DATE: December 6, 1994
CC APPLICATION NUMBER: 08/226,248
CC FILING DATE: April 8, 1994
CC APPLICATION NUMBER: 07/938,990
CC FILING DATE: September 1, 1992
CC APPLICATION NUMBER: 08/789,301
CC FILING DATE: January 15, 1993
CC APPLICATION NUMBER: 07/938,990
CC FILING DATE: September 1, 1992
CC APPLICATION NUMBER: 08/US93/00134
CC FILING DATE: January 15, 1993
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Darlene A. Vanstone
CC REGISTRATION NUMBER: 35,729
CC REFERENCE/DOCKET NUMBER: 025.6 US (IM-028CP2)
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (617) 466-6000
CC TELEFAX: (617) 466-6040
CC INFORMATION FOR SEQ ID NO: 62:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 60 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC FRAGMENT TYPE: internal
CC SEQUENCE 60 AA: 6644 MW: 19464 CN:
SQ
Query Match 100.0% Score 64: DB 10: Length 60:
Best Local Similarity 100.0% Pred. No. 2,93e+00:
Matches 9: Conservative 0: Mismatches 0: Indels 0: Gaps 0:
Db 49 FIKRVSNVI 57
QY : FIKRVSNVI 9

```

```

RESULT 12
ID US-08-226-248A-62 STANDARD: PRT: 60 AA.
XX
AC xxxxxx
XX
DT
XX
DE Sequence 62, Application: US/08226248A
XX
CC Sequence 62, Application US/08226248A
CC GENERAL INFORMATION:
CC APPLICANT: Griffith, Irwin J.:
CC APPLICANT: Pollock, Joanne:
CC APPLICANT: Bond, Julian F.:
CC APPLICANT: Garman, Richard D.:
CC APPLICANT: Kuo, Mei-Chang:
CC APPLICANT: Yeung, Siu-mei H.:
CC APPLICANT: Brauer, Andrew:
CC APPLICANT: Exley, Mark A.:
CC APPLICANT: Powers, Steven P.:
CC TITLE OF INVENTION: Allergenic Proteins And Peptides From
CC TITLE OF INVENTION: Japanese Cedar Pollen
CC NUMBER OF SEQUENCES: 201
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Immunologic Pharmaceutical Corporation, Inc.
CC STREET: 610 Lincoln St
CC CITY: Waltham
CC STATE: MA
CC COUNTRY: USA
CC ZIP: 02154
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patent Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/226.248A
CC FILING DATE: April 8, 1994
CC CLASSIFICATION: 424
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 07/938,990
CC FILING DATE: September 1, 1992
CC APPLICATION NUMBER: 08/789,301
CC FILING DATE: January 15, 1993
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Darlene A. Vanstone
CC REGISTRATION NUMBER: 35,729
CC REFERENCE/DOCKET NUMBER: 025.6 US (IM-028CP)
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (617) 466-6000
CC TELEFAX: (617) 466-6040
CC INFORMATION FOR SEQ ID NO: 62:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 60 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC FRAGMENT TYPE: internal
CC SEQUENCE 60 AA: 6644 MW: 19464 CN:
SQ
Query Match 100.0% Score 64: DB 7: Length 60:
Best Local Similarity 100.0% Pred. No. 2,93e+00:
Matches 9: Conservative 0: Mismatches 0: Indels 0: Gaps 0:
Db 49 FIKRVSNVI 57
QY 1 FIKRVSNVI 9
RESULT 13
ID US-08-467-697-62 STANDARD: PRT: 60 AA.

```

```
xx      xxxxxx
AC
XX
DT
XX Sequence 62, Application US/08457697
XX
CC Sequence 62, Application US/08457697
CC GENERAL INFORMATION:
CC APPLICANT: Griffith, Irwin J.
CC APPLICANT: Pollock, Joanne
CC APPLICANT: Bond, Julian F.
CC APPLICANT: Garmar, Richard D.
CC APPLICANT: Kuo, Mei-Chang
CC APPLICANT: Yeung, Siu-mei H.
CC APPLICANT: Brauer, Andrew
CC APPLICANT: Exley, Mark A.
CC APPLICANT: Powers, Steven P.
CC TITLE OF INVENTION: Allergenic Proteins And Peptides From
CC TITLE OF INVENTION: Japanese Cedar Pollen
CC NUMBER OF SEQUENCES: 261
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Immunologic Pharmaceutical Corporation, Inc.
CC STREET: 610 Lincoln St.
CC CITY: Waltham
CC STATE: MA
CC COUNTRY: USA
CC ZIP: 02154
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patent In Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/457,697
CC FILING DATE: June 6, 1995
CC CLASSIFICATION: 424
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/350,225
CC FILING DATE: December 6, 1994
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Jane E. Remillard
CC REGISTRATION NUMBER: 38,872
CC REFERENCE/DOCKET NUMBER: 025.6 USD4 (IM-028CPD4)
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (517) 227-7400
CC TELEFAX: (517) 227-5941
CC INFORMATION FOR SEQ ID NO: 62:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 60 amino acids
CC MOLECULE TYPE: linear
CC FRAGMENT TYPE: internal
CC SEQUENCE 60 AA: 6644 MW: 19464 CN:

Query Match 100.0% Score 64: DB 10: Length 60:
Best Local Similarity 100.0% Pred. No. 2,93e-00:
Matches 9: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

Db 43 FIKRVSNVI 57
QY 1 FIKRVSNVI 9

RESULT 14
ID US-08-457-023-62 STANDARD: PRT: 60 AA.
AC xxxxxx
XX
DT
XX Sequence 62, Application US/08457697
XX
CC Sequence 62, Application US/08457697
CC GENERAL INFORMATION:
CC APPLICANT: Griffith, Irwin J.
CC APPLICANT: Pollock, Joanne
CC APPLICANT: Bond, Julian F.
CC APPLICANT: Garmar, Richard D.
CC APPLICANT: Kuo, Mei-Chang
CC APPLICANT: Yeung, Siu-mei H.
CC APPLICANT: Brauer, Andrew
CC APPLICANT: Exley, Mark A.
CC APPLICANT: Powers, Steven P.
CC TITLE OF INVENTION: Allergenic Proteins And Peptides From
CC TITLE OF INVENTION: Japanese Cedar Pollen
CC NUMBER OF SEQUENCES: 261
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Immunologic Pharmaceutical Corporation, Inc.
CC STREET: 610 Lincoln St.
CC CITY: Waltham
CC STATE: MA
CC COUNTRY: USA
CC ZIP: 02154
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patent In Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/457,697
CC FILING DATE: June 6, 1995
CC CLASSIFICATION: 424
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/350,225
CC FILING DATE: December 6, 1994
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Jane E. Remillard
CC REGISTRATION NUMBER: 38,872
CC REFERENCE/DOCKET NUMBER: 025.6 USD4 (IM-028CPD4)
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (517) 227-7400
CC TELEFAX: (517) 227-5941
CC INFORMATION FOR SEQ ID NO: 62:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 60 amino acids
CC MOLECULE TYPE: linear
CC FRAGMENT TYPE: internal
CC SEQUENCE 60 AA: 6644 MW: 19464 CN:

Query Match 100.0% Score 64: DB 10: Length 60:
Best Local Similarity 100.0% Pred. No. 2,93e-00:
Matches 9: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

Db 43 FIKRVSNVI 57
QY 1 FIKRVSNVI 9

RESULT 14
ID US-08-457-023-62 STANDARD: PRT: 60 AA.
AC xxxxxx
XX
DT
XX Sequence 62, Application US/08457697
XX
CC Sequence 62, Application US/08457697
CC GENERAL INFORMATION:
CC APPLICANT: Griffith, Irwin J.
CC APPLICANT: Pollock, Joanne
CC APPLICANT: Bond, Julian F.
```

```
xx      xxxxxx
AC
XX
DT
XX Sequence 62, Application US/ 457023
XX
CC Sequence 62, Application US/ 457023
CC GENERAL INFORMATION:
CC APPLICANT: Griffith, Irwin J.
CC APPLICANT: Pollock, Joanne
CC APPLICANT: Bond, Julian F.
CC APPLICANT: Garmar, Richard D.
CC APPLICANT: Kuo, Mei-Chang
CC APPLICANT: Yeung, Siu-mei H.
CC APPLICANT: Brauer, Andrew
CC APPLICANT: Exley, Mark A.
CC APPLICANT: Powers, Steven P.
CC TITLE OF INVENTION: Allergenic Proteins And Peptides From
CC TITLE OF INVENTION: Japanese Cedar Pollen
CC NUMBER OF SEQUENCES: 261
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Immunologic Pharmaceutical Corporation, Inc.
CC STREET: 610 Lincoln St.
CC CITY: Waltham
CC STATE: MA
CC COUNTRY: USA
CC ZIP: 02154
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patent In Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/457,023
CC FILING DATE: June 6, 1995
CC CLASSIFICATION: 424
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/350,225
CC FILING DATE: December 6, 1994
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Jane E. Remillard
CC REGISTRATION NUMBER: 38,872
CC REFERENCE/DOCKET NUMBER: 025.6 USD2 (IM-028CPD2)
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (617) 227-7400
CC TELEFAX: (617) 227-5941
CC INFORMATION FOR SEQ ID NO: 62:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 60 amino acids
CC MOLECULE TYPE: linear
CC FRAGMENT TYPE: internal
CC SEQUENCE 60 AA: 6644 MW: 19464 CN:

Query Match 100.0% Score 64: DB 10: Length 60:
Best Local Similarity 100.0% Pred. No. 2,93e-00:
Matches 9: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

Db 43 FIKRVSNVI 57
QY 1 FIKRVSNVI 9

RESULT 15
ID US-08-350-225-62 STANDARD: PRT: 60 AA.
AC xxxxxx
XX
DT
XX
DE
XX Sequence 62, Application US/08450225
XX
CC Sequence 62, Application US/08350225
CC GENERAL INFORMATION:
CC APPLICANT: Griffith, Irwin J.
CC APPLICANT: Pollock, Joanne
CC APPLICANT: Bond, Julian F.
```



\*\*\*\*\*  
M B E R E E L  
\*\*\*\*\*  
(TM)

Release 3.2A John F. Collins, Bioconforming Research Unit.  
Copyright (C) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run On: Mon Jun 19 16:06:27 2000; MasPar time 5.08 Seconds  
Tatular output not generated. 83.618 Million cell updates/sec

Title: >US-09-142-524A-7  
Description: (1-9) from US09142524A.pep  
Perfect Score: 64  
Sequence: 1 FIKRVSNV: 9

Scoring table: BAW 150  
Gap 15

Searched: 142080 seqs. 47:72406 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: pir62  
1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 23.798 Variance 28.919 scale 0.823

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Query Match | Length | ID   | Description                 | Pred. No. |
|------------|-------------|--------|------|-----------------------------|-----------|
| 1          | 64          | 100.0  | 374  | major allergen Cry j 1      | 8.03e-03  |
| 2          | 64          | 100.0  | 374  | major allergen Cry j 1      | 8.03e-03  |
| 3          | 67          | 89.1   | 326  | peptidase - Asper           | 3.16e-01  |
| 4          | 52          | 79.7   | 720  | hypothetical protein        | 5.98e-00  |
| 5          | 50          | 78.1   | 256  | beta-lactamase (EC 3.1.1.1) | 9.56e-00  |
| 6          | 50          | 78.1   | 317  | endo-glucanase - Therm      | 9.56e-00  |
| 7          | 50          | 78.1   | 473  | hypothetical protein        | 9.56e-00  |
| 8          | 50          | 78.1   | 479  | hypothetical protein        | 9.56e-00  |
| 9          | 49          | 76.6   | 190  | transcription regulat       | 1.52e-01  |
| 10         | 49          | 76.6   | 227  | beta-lactamase (EC 3.1.1.1) | 1.52e-01  |
| 11         | 49          | 76.6   | 227  | beta-lactamase (EC 3.1.1.1) | 1.52e-01  |
| 12         | 49          | 76.6   | 379  | probable membrane pro       | 1.52e-01  |
| 13         | 49          | 76.6   | 379  | probable membrane pro       | 1.52e-01  |
| 14         | 49          | 76.6   | 381  | probable membrane pro       | 1.52e-01  |
| 15         | 49          | 76.6   | 432  | hypothetical protein        | 1.52e-01  |
| 16         | 49          | 76.6   | 520  | cysteine--trna ligase       | 1.52e-01  |
| 17         | 48          | 75.0   | 487  | carbamoyl-phosphate s       | 2.39e-01  |
| 18         | 48          | 75.0   | 601  | probable glycosyl hyd       | 2.39e-01  |
| 19         | 48          | 75.0   | 4725 | dynein heavy chain, c       | 2.39e-01  |
| 20         | 47          | 73.4   | 234  | hypothetical protein        | 3.75e-01  |
| 21         | 47          | 73.4   | 410  | 3-phosphoglycerate ki       | 3.75e-01  |
| 22         | 47          | 73.4   | 511  | dihydrolipicamide S-su      | 3.75e-01  |
| 23         | 47          | 73.4   | 537  | probable flagella-rel       | 3.75e-01  |

|    |    |      |      |   |        |                       |          |
|----|----|------|------|---|--------|-----------------------|----------|
| 24 | 47 | 73.4 | 718  | 2 | C72285 | hypothetical protein  | 3.75e-01 |
| 25 | 47 | 73.4 | 899  | 2 | S49534 | hypothetical protein  | 3.75e-01 |
| 26 | 47 | 73.4 | 1462 | 1 | S32437 | pol polyprotein - Vol | 3.75e-01 |
| 27 | 46 | 71.9 | 143  | 2 | A69115 | hypothetical protein  | 5.92e-01 |
| 28 | 46 | 71.9 | 317  | 2 | G70356 | thymidylate synthase  | 5.92e-01 |
| 29 | 46 | 71.9 | 500  | 1 | Q08E08 | HHLF6 protein - human | 5.82e-01 |
| 30 | 46 | 71.9 | 528  | 2 | A37878 | glucuronosyltransfera | 5.82e-01 |
| 31 | 46 | 71.9 | 528  | 2 | JN0619 | JAP-glucuronosyltrans | 5.82e-01 |
| 32 | 46 | 71.9 | 530  | 2 | A36276 | glucuronosyltransfera | 5.82e-01 |
| 33 | 46 | 71.9 | 530  | 2 | S07390 | glucuronosyltransfera | 5.82e-01 |
| 34 | 46 | 71.9 | 530  | 2 | A42667 | glucuronosyltransfera | 5.82e-01 |
| 35 | 46 | 71.9 | 530  | 2 | A25064 | glucuronosyltransfera | 5.82e-01 |
| 36 | 46 | 71.9 | 532  | 2 | T16752 | hypothetical protein  | 5.82e-01 |
| 37 | 46 | 71.9 | 545  | 1 | JX0225 | cytochrome P450 CYP19 | 5.82e-01 |
| 38 | 46 | 71.9 | 585  | 2 | S73325 | stage II sporulation  | 5.82e-01 |
| 39 | 46 | 71.9 | 585  | 2 | S71048 | stage II sporulation  | 5.82e-01 |
| 40 | 46 | 71.9 | 774  | 2 | T14555 | DNA polymerase hpolo  | 5.82e-01 |
| 41 | 46 | 71.9 | 840  | 1 | I40509 | endopeptidase Clp AIP | 5.82e-01 |
| 42 | 46 | 71.9 | 847  | 2 | A48228 | beta-N-acetylhexosami | 5.82e-01 |
| 43 | 46 | 71.9 | 857  | 1 | D35905 | endopeptidase Clp AIP | 5.82e-01 |
| 44 | 46 | 71.9 | 1006 | 2 | C70445 | ATPase subunit of ATP | 5.82e-01 |
| 45 | 46 | 71.9 | 1274 | 2 | T04018 | hypothetical protein  | 5.82e-01 |

ALIGNMENTS

RESULT: 1  
ENTRY: JC2124 \*type complete  
TITLE: major allergen Cry j 1 precursor (clone pCC1-15) - Japanese cedar  
ORGANISM: \*formal\_name Cryptomeria japonica \*common\_name Japanese cedar  
DATE: 14-Jul-1994 \*sequence\_revision 14-Jul-1994 \*text\_change 26-Aug-1999  
ACCESSIONS: JC21124  
REFERENCE: JC21123  
\*authors: Sone, T.; Komiyama, N.; Shimizu, K.; Kusakabe, T.; Murikubo, K.; Kino, K.  
\*journal: Biochem. Biophys. Res. Commun. (1994) 199:519-525  
\*title: Cloning and sequencing of cDNA coding for Cry j 1, a major allergen of Japanese cedar pollen.  
\*cross-references: MIM:9416334  
\*accession: JC2124  
\*molecule\_type: mRNA  
\*residues: 1-374 \*label: SGN  
\*cross-references: G0125459; N0149494; P01410687; P01410684  
\*experimental\_data: Source: pollen.  
\*note: The authors described a cDNA clone encoding a protein for residue 27.

CLASSIFICATION: \*superfamily: peptidase lyase (A1)  
KEYWORDS: glycoprotein; pollen  
FEATURES:  
1-21  
22-374  
158,191,293,354 \*length: 374 \*molecular\_weight: 40702 \*checksum: 3592  
\*domain signal: sequence status predicted \*label: SGN  
\*product: major allergen cry j 1 (clone pCC1-15) \*status: predicted \*label: MAT  
\*binding site: carbohydrate (Asn) (covalent) \*status: predicted  
SUMMARY: \*length: 374 \*molecular\_weight: 40702 \*checksum: 3592  
Query Match: 100.0%; Score: 64; DB: 2; Length: 374;  
Best Local Similarity: 100.0%; Frac. No.: 8.03e-03;  
Matches: 9; Conservative: 1; Mismatches: 0; Indels: 0; Gaps: 0;

DB: 130 FIKRVSNV: 138  
QY: 1 FIKRVSNV: 9  
RESULT: 2  
ENTRY: JC2123 \*type complete  
TITLE: major allergen Cry j 1 precursor (clone pCC1-2-2) - Japanese cedar  
ORGANISM: \*formal\_name Cryptomeria japonica \*common\_name Japanese cedar  
DATE: 14-Jul-1994 \*sequence\_revision 14-Jul-1994 \*text\_change

```

26-Aug-1999
ACCESSIONS JC2123: PC2065
REFERENCE   JC2123
AUTHORS     Suet, T.; Komiyama, N.; Shimizu, K.; Kusakabe, T.; Morikubo, K.; Kiro, K.
JOURNAL     Biochem. Biophys. Res. Commun. (1994) 199:619-625
TITLE       Cloning and sequencing of cDNA coding for Cry 1A, a major allergen of Japanese cedar pollen.
CROSS-REFERENCES MIM:34103234
ACCESSION   JC2123
MOLECULE_TYPE DNA
RESIDUES    1-374 **label SON
CROSS-REFERENCES GB:D26544; NID:q493631; PID:d1006086; PID:g493632
EXPERIMENTAL_SOURCE pollen
ACCESSION   PC2065
MOLECULE_TYPE protein
RESIDUES    22-53:58-81:219-232:236-258:259-357:365-372 **label SC2
NOTE        the authors described carbohydrate binding site for residue 279
CLASSIFICATION *superfamily pectate lyase LA159
KEYWORDS       glycoprotein; pollen.
FEATURE
1-2:          *domain signal sequence *status predicted *label SIGN
22-374        *product major allergen Cry 1A (clone PCC1-2-2); *status predicted *label MAIN
159-191:233,354 *binding_site carbohydrate (Asn) (covalent) *status predicted
SUMMARY        length 374 *molecular_weight 40645 *checksum 2920
Query Match    100.0% Score 64 DB 2 Length 374
Best Local Similarity 100.0% Pred. No. 8.03e-03%
Matches 97 Conservative 0 Mismatches 0 Indels 0 Gaps 0
DB 110 YKRVSNV1 138
QY 1 FIKRVSNV1 9
RESULT 3
ENTRY     SS1509 **type complete
TITLE     pectate lyase - Aspergillus sp.
ORGANISM  *formal_name Aspergillus sp.
DATE      15-Jun-1995 *sequence_revision 15-Apr-1995 *text_change 19-Apr-1996
ACCESSION SS1509
REFERENCE  RM: M.C. Whitford, M.F. Cleveland, E. Douthett, R.A. Gurr, Genet. (1996) 27:142-149
AUTHORS    Gurr, Genet. (1996) 27:142-149
JOURNAL    Sequence analysis of the Aspergillus nidulans pectate lyase beta gene and evidence for binding of promoter regions to CRE, a regulator of carbon catabolite repression.
CROSS-REFERENCES MIM:D5308536
ACCESSION  SS1509
STATUS     preliminary
MOLECULE_TYPE DNA
RESIDUES    1-326 **label HOM
SUMMARY    length 326 *molecular_weight 34565 *checksum 3638
Query Match    89.1% Score 57 DB 2 Length 326
Best Local Similarity 77.8% Pred. No. 3.16e-01%
Matches 77 Conservative 2 Mismatches 0 Indels 0 Gaps 0
DB 110 YKRVSNV1 118
QY 1 FIKRVSNV1 9
RESULT 4
ENTRY     SC0920 **type complete
TITLE     hypothetical protein YKR192w - yeast (Saccharomyces cerevisiae)
ORGANISM  *formal_name Saccharomyces cerevisiae
ALTERNATE_NAMES YK9646.04
ORGANISM

```

```

10-Feb-1995 *sequence_revision 12-May-1995 *text_change 21-Nov-1997
ACCESSIONS S50920
REFERENCE   S50917
AUTHORS     Pearson, D.; Rowan, S.
JOURNAL     Submitted to the EMBL Data Library, January 1995
TITLE       Cloning and sequencing of cDNA coding for Cry 1A, a major allergen of Japanese cedar pollen.
CROSS-REFERENCES MIM:34103234
ACCESSION   S50920
MOLECULE_TYPE DNA
RESIDUES    1-720 **label PE4
CROSS-REFERENCES EMBL:Z47915; NID:q642280; PID:q642284; MIM: YK9646.04
GENETICS
MAP POSITION 138
SUMMARY    length 720 *molecular_weight 82131 *checksum 646
Query Match    79.7% Score 53 DB 2 Length 720
Best Local Similarity 55.6% Pred. No. 5.98e-06%
Matches 57 Conservative 4 Mismatches 0 Indels 0 Gaps 0
DB 351 YKRVSNV1 359
QY 1 FIKRVSNV1 9
RESULT 5
ENTRY     A32017 **type complete
TITLE     beta-lactamase (EC 3.5.1.6) II precursor - Bacillus cereus (strain 5/B/5)
ORGANISM  *formal_name Bacillus cereus
DATE      07-Jun-1990 *sequence_revision 07-Jun-1990 *text_change 22-Jun-1995
ACCESSIONS A32017
REFERENCE  Lin, H.M.; Penev, J.; Shaw, K.W.; J. Bacteriol. (1988) 170:2873-2878
AUTHORS    Cloning, nucleotide sequence, and expression of the bacillus cereus 5/B/6 beta-lactamase II structural gene.
CROSS-REFERENCES MIM:D8227875
ACCESSION  A32017
MOLECULE_TYPE DNA
RESIDUES    1-256 **label IM
CROSS-REFERENCES GB:M19530; MIM:q143127; PID:AAA22562; PID:q143129
CLASSIFICATION *superfamily beta-lactamase II
KEYWORDS     antibiotic resistance; hydrolase
FEATURE
1-256       *domain signal sequence *status predicted *label SIGN
SUMMARY    length 256 *molecular_weight 28148 *checksum 121
Query Match    74.3% Score 60 DB 2 Length 256
Best Local Similarity 64.7% Pred. No. 9.56e-09%
Matches 59 Conservative 2 Mismatches 1 Indels 0 Gaps 0
DB 104 FIKRVTDV1 112
QY 1 FIKRVSNV1 9
RESULT 6
ENTRY     B72216 **type complete
TITLE     endoglucanase - Thermotoga maritima (strain MSB8)
ORGANISM  *formal_name Thermotoga maritima
DATE      11-Jun-1999 *sequence_revision 11-Jun-1999 *text_change 11-Jun-1999
ACCESSIONS B72216
REFERENCE   A72200
AUTHORS     Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey, E.K.; Peterson, J.D.; Neeson, W.C.; Ketchum, K.A.; McDonald, L.; Utterback, T.R.; Mailek, J.A.; Linher, M.D.; Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.; Heidelberg, J.; Sutton, G.G.; Fleischmann, R.D.; White, O.; Salzberg, S.L.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
JOURNAL     Nature (1999) 399:323-329

```

```

#title      Evidence for lateral gene transfer between Archaea and
#cross-references MUID:9287316
#accession   B72216
#status      preliminary
#molecule_type DNA
#residues 1-317 ##label ARN
#cross-references GB:AE001613; GB:AE000512; NID:q4982323;
#experimental_source strain: MSB8

GENETICS
#gene
#summary
  Query Match      78.1%; Score 50; DB 2; Length 317;
  Best Local Similarity 66.7%; Pred. No. 9.56e+00;
  Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

DB 72 FFKRVDEVI 90
  1111111
QY 1 FIKRVSNVI 9

RESULT 7
ENTRY      370541      ##type complete
TITLE      hypothetical protein RV1576c - Mycobacterium tuberculosis
           (strain H37RV)
ORGANISM   #format_name Mycobacterium tuberculosis
DATE       17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change
           17-Jul-1998

ACCESSIONS
REFERENCE   B70500
#authors   Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher,
           C.; Harris, D.; Gordon, S.V.; Eigmeier, K.; Gas, S.; Barry
           III, C.E.; Tekala, F.; Badcock, K.; Basham, D.; Brown, D.;
           Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.;
           Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.;
           Hornsby, T.; Jagels, K.; Krogh, A.; McLean, J.; Mule, S.;
           Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.;
           Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.;
           Skellton, S.; Squares, S.; Squires, R.; Sulston, J.E.;
           Taylor, K.; Whitehead, S.; Barrrell, B.G.
#journal   Nature (1998) 393:537-544
#title     Deciphering the biology of Mycobacterium tuberculosis from
           the complete genome sequence.
#cross-references MUID:92895987
#accession   B70541
#status      preliminary; nucleic acid sequence not shown;
           translation not shown;
           ##molecule_type DNA
           ##residues 1-479 ##label COL
           ##cross-references GB:Z80225; GB:AL123456; NID:q3242265; PID:q4981354;
           PID:q35069;
           ##experimental_source strain H37RV

GENETICS
#gene
#summary
  Query Match      78.1%; Score 50; DB 2; Length 479;
  Best Local Similarity 66.7%; Pred. No. 9.56e+00;
  Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

DB 154 FVKRVSNPV 162
  1111111
QY 1 FIKRVSNVI 9

RESULT 9
ENTRY      B72330      ##type complete
TITLE      transcription regulator, ter family - Thermotoga maritima
           (strain MSB6)
ORGANISM   #format_name Thermotoga maritima
DATE       11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change
           20-Aug-1999

ACCESSIONS
REFERENCE   A72330
#authors   Nelson, K.R.; Clayton, K.A.; Gill, S.R.; Binkley, M.E.; Dodson,
           R.J.; Holt, D.B.; Hickey, E.K.; Peterson, J.D.; Nelson,
           W.C.; Ketchum, K.A.; McDonald, L.; Utterback, L.; Malek,
           J.A.; Linher, K.D.; Garrett, M.M.; Stewart, A.M.; Cotton,
           M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.;
           Heidelberg, J.; Sutton, G.G.; Fleischmann, R.D.; White, O.;
           Salzberg, S.L.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
#journal   Nature (1999) 399:323-329
#title     Evidence for lateral gene transfer between Archaea and
           Bacteria from genome sequence of Thermotoga maritima.
#cross-references MUID:9287316
#accession   B72330
#status      preliminary
#molecule_type DNA
#residues 1-190 ##label ARN
#cross-references GB:AE001749; GB:AE000512; NID:q4981346; PID:q4981354;
           TIGR:TM0823
           ##experimental_source strain MSB8

GENETICS
#gene
#summary
  Query Match      76.6%; Score 49; DB 2; Length 190;
  Best Local Similarity 66.7%; Pred. No. 1.52e+01;
  Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

DB 148 FVKRVSNPV 156
  1111111
QY 1 FIKRVSNVI 9

RESULT 8
ENTRY      F70965
TITLE      hypothetical protein RV2650c - Mycobacterium tuberculosis
           (strain H37RV)
ORGANISM   #format_name Mycobacterium tuberculosis
DATE       17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change

```

```

17-Jul-1998
F70965
A70500
#authors   Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher,
           C.; Harris, D.; Gordon, S.V.; Eigmeier, K.; Gas, S.; Barry
           III, C.E.; Tekala, F.; Badcock, K.; Basham, D.; Brown, D.;
           Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.;
           Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.;
           Hornsby, T.; Jagels, K.; Krogh, A.; McLean, J.; Mule, S.;
           Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.;
           Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.;
           Skellton, S.; Squares, S.; Squires, R.; Sulston, J.E.;
           Taylor, K.; Whitehead, S.; Barrrell, B.G.
#journal   Nature (1998) 393:537-544
#title     Deciphering the biology of Mycobacterium tuberculosis from
           the complete genome sequence.
#cross-references MUID:92895987
#accession   F70965
#status      preliminary; nucleic acid sequence not shown;
           translation not shown;
           ##molecule_type DNA
           ##residues 1-479 ##label COL
           ##cross-references GB:Z80225; GB:AL123456; NID:q3242265; PID:q4981354;
           PID:q35069;
           ##experimental_source strain H37RV

GENETICS
#gene
#summary
  Query Match      78.1%; Score 50; DB 2; Length 479;
  Best Local Similarity 66.7%; Pred. No. 9.56e+00;
  Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

DB 154 FVKRVSNPV 162
  1111111
QY 1 FIKRVSNVI 9

RESULT 9
ENTRY      B72330      ##type complete
TITLE      transcription regulator, ter family - Thermotoga maritima
           (strain MSB6)
ORGANISM   #format_name Thermotoga maritima
DATE       11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change
           20-Aug-1999

ACCESSIONS
REFERENCE   A72330
#authors   Nelson, K.R.; Clayton, K.A.; Gill, S.R.; Binkley, M.E.; Dodson,
           R.J.; Holt, D.B.; Hickey, E.K.; Peterson, J.D.; Nelson,
           W.C.; Ketchum, K.A.; McDonald, L.; Utterback, L.; Malek,
           J.A.; Linher, K.D.; Garrett, M.M.; Stewart, A.M.; Cotton,
           M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.;
           Heidelberg, J.; Sutton, G.G.; Fleischmann, R.D.; White, O.;
           Salzberg, S.L.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
#journal   Nature (1999) 399:323-329
#title     Evidence for lateral gene transfer between Archaea and
           Bacteria from genome sequence of Thermotoga maritima.
#cross-references MUID:9287316
#accession   B72330
#status      preliminary
#molecule_type DNA
#residues 1-190 ##label ARN
#cross-references GB:AE001749; GB:AE000512; NID:q4981346; PID:q4981354;
           TIGR:TM0823
           ##experimental_source strain MSB8

GENETICS
#gene
#summary
  Query Match      76.6%; Score 49; DB 2; Length 190;
  Best Local Similarity 66.7%; Pred. No. 1.52e+01;
  Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

DB 148 FVKRVSNPV 156
  1111111
QY 1 FIKRVSNVI 9

RESULT 8
ENTRY      F70965
TITLE      hypothetical protein RV2650c - Mycobacterium tuberculosis
           (strain H37RV)
ORGANISM   #format_name Mycobacterium tuberculosis
DATE       17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change

```

```

#accession: A91344
#molecule_type: protein
#keyresidues 31-183;187-210;214-257 ##label AMB
#experimental_source strain 569/H/9
COMMENT beta-lactamase II from Staphylococcus aureus strain 569/H/9. This
pseudomonas maltophilia are classified together as class B
beta-lactamases, which hydrolyze cephalosporins and penicillins.
beta-lactamase II binds two Zn(II) ions per molecule. Zn at the
higher affinity site is necessary for activity of the enzyme.
Zinc at the lower affinity site (not specified) increases
hydrolysis of cephalosporin C but not of benzylpenicillin.
KEYWORDS antibiotic resistance; hydrolase; metalloprotein; zinc
FEATURE
1-30 #domain signal sequence #status predicted #db: SSG
31-257 #product beta-lactamase II #status experimental #db:
MPTN
67 #active_site Glu #status predicted\
116,118,198,240 #binding_site zinc, high affinity (his, his, lys, his)
#status experimental
SUMMARY #length 257 #molecular_weight 28042 #checksum 428
Query Match 76.6% Score 49: 58 1: Length 257:
Best Local similarity 66.7% Pred No 1,528-01:
Matches 6: Conservative 2: Mismatches 1: Indels 0: Gaps 0:
Db 105 F0KRVTOVI 113
| | | | |
QY 1 FIKRVSNVI 9
RESULT 12
ENTRY
TITLE
ALTERNATE_NAMES
ORGANISM
DATE
ACCESSIONS
REFERENCE
#authors
#submitters
#description
#accession: S62275
#molecule_type: DNA
#keyresidues 1-379 ##label MGR
#cross-references EMBL:D45603; NID:q871957; PIDN:BAA06057.1
#cross-references PDB:d009817; PID:q836593; MIPS:YF1624
S62230
#authors
#submitters
#accession: S62275
#molecule_type: DNA
#keyresidues 1-379 ##label MGR
#cross-references EMBL:D45603; NID:q871957; PIDN:BAA06057.1
#cross-references PDB:d009817; PID:q836593; MIPS:YF1624
GENETICS
#gene
#cross-references SGO:SC 1832; MIPS:YF1062W
#map_position 6L
CLASSIFICATION
KEYWORDS
FEATURE
75-91 #domain transmembrane #status predicted #label TM1\
236-252 #domain transmembrane #status predicted #label TM2\
236-272 #domain transmembrane #status predicted #label TM3\
SUMMARY #length 379 #molecular_weight 45312 #checksum 142
Query Match 76.6% Score 49: DB 2: Length 379:

```



```

254-273      *domain transmembrane *status predicted *label TM4
SUMMARY      *length 379 *molecular-weight 45165 *checksum 1211
Query Match  76.6% Score 49 DB 2 Length 379
Best Local Similarity 55.6% F ad NC 1.52e-01
Matches      5: Conservative Mismatches 0: Indels 0: Gaps 0:
DB 229 FLKRSNIF 237
      :|||||:
QY 1 FIKRVSNVI 9

RESULT 13
ENTRY   S45187 *type complete
TITLE   probable membrane protein YKL219W homolog YBR302C - yeast
        (Saccharomyces cerevisiae)
ALTERNATE_NAMES  hypothetical protein YBR2121: hypothetical protein YK4987.03:
        hypothetical protein YML132:
ORGANISM  *formal_name Saccharomyces cerevisiae
DATE      26-Aug-1994 *sequence_revision 09-Sep-1994 *text_change
ACCESSIONS  S45187: S46186: S55818: S58196
REFERENCE   Feldmann, H.; Mannhaupt, G.; Schwarziolse, C.; Vetter, I.
            Submitted to the Protein Sequence Database, August 1994
AUTHORS     Feldmann, H.; Mannhaupt, G.; Schwarziolse, C.; Vetter, I.
SUBMISSION  Submitted to the Protein Sequence Database, August 1994
ACCESSION   S45187
            *molecule_type DNA
            *residues 1-47 *label PEL
            *cross-references EMBL:236170: MIPS:YBR302C
            *experimental_source strain S288C
            *genetics C02
REFERENCE   S46181
            Federmann, M.; Potier, S.; Souciet, J.L.
            Submitted to the Protein Sequence Database, August 1994
AUTHORS     Federmann, M.; Potier, S.; Souciet, J.L.
SUBMISSION  Submitted to the Protein Sequence Database, August 1994
ACCESSION   S46181
            *molecule_type DNA
            *residues 30-379 *label FEU
            *cross-references EMBL:236170: MIPS:YBR302C
            *experimental_source strain S288C
            *genetics C02
REFERENCE   S55813
            Federmann, M.; Charbonnel, L.; de Montigny, J.; Bloch, J.C.;
            Potier, S.; Souciet, J.L.
            Yeast (1995) 11:667-672
            Sequence of a 9.8 kb segment of yeast chromosome II: including
            the three genes of the MAL3 locus and three unidentified
            open reading frames.
            *cross-references KJ039609/939
            *accession S55813
            *status translation not shown
            *molecule_type DNA
            *residues 30-379 *label FEU
            *cross-references EMBL:236170:
            *genetics C02
REFERENCE   S58194
            Rowman, S.
            Submitted to the EMBL Data Library, July 1995
            *accession S58194
            *molecule_type DNA
            *residues 1-379 *label BOW
            *cross-references EMBL:250178: NID:G927528: PIDN:CAA90551.1:
            PID:G927530: MIPS:YML132W
            *experimental_source strain AB972
            *genetics C02
            *cross-references MIPS:YML132W: SGD:S0004601
            *map_position 13L
            *classification *superfamily conserved hypothetical protein: YKL219W
            *keywords transmembrane protein
FEATURE     46-62      *domain transmembrane *status predicted *label TM1\
            65-96      *domain transmembrane *status predicted *label TM2\
            233-253    *domain transmembrane *status predicted *label TM3\

```

```

254-273      *domain transmembrane *status predicted *label TM4
SUMMARY      *length 379 *molecular-weight 45165 *checksum 1211
Query Match  76.6% Score 49 DB 2 Length 379
Best Local Similarity 55.6% F ad NC 1.52e-01
Matches      5: Conservative Mismatches 0: Indels 0: Gaps 0:
DB 229 FLKRSNIF 237
      :|||||:
QY 1 FIKRVSNVI 9

RESULT 14
ENTRY   S63322 *type complete
TITLE   probable membrane protein YNL336W - yeast (Saccharomyces
        cerevisiae)
ALTERNATE_NAMES  hypothetical protein N0275
ORGANISM  *formal_name Saccharomyces cerevisiae
DATE      27-Apr-1996 *sequence_revision 03-May-1996 *text_change
ACCESSIONS  S63322
REFERENCE   S63317
            Obermaier, B.; Pittavandi, E.; Rucke, M.
            Submitted to the Protein Sequence Database, April 1996
SUBMISSION  Submitted to the Protein Sequence Database, April 1996
ACCESSION   S63322
            *molecule_type DNA
            *residues 1-381 *label OBE
            *cross-references EMBL:271611: NID:q1302452: PIDN:CAA94270.1:
            PID:q23973: PID:q1302453: MIPS:YNL336W
            *experimental_source strain S288C
            *genetics SGD:COS1
            *cross-references SGD:S000540: MIPS:YNL336W
            *map_position 14L
            *classification *superfamily conserved hypothetical protein YKL219W
            *keywords transmembrane protein
            *feature
            74-90      *domain transmembrane *status predicted *label TM1\
            236-252    *domain transmembrane *status predicted *label TM2\
            256-272    *domain transmembrane *status predicted *label TM3\
SUMMARY      *length 381 *molecular-weight 45288 *checksum 6842
Query Match  76.6% Score 49 DB 2 Length 381
Best Local Similarity 55.6% F ad NC 1.52e-01
Matches      5: Conservative Mismatches 0: Indels 0: Gaps 0:
DB 229 FLKRSNIF 237
      :|||||:
QY 1 FIKRVSNVI 9

RESULT 15
ENTRY   G64403 *type complete
TITLE   hypothetical protein MJ0831 - Methanococcus jannaschii
ORGANISM  *formal_name Methanococcus jannaschii
DATE      13-Sep-1995 *sequence_revision 13-Sep-1995 *text_change
ACCESSIONS  G64403
REFERENCE   A54300
            Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann,
            R.D.; Sutton, G.G.; Blake, J.A.; Fitzgerald, L.M.; Clayton,
            R.A.; Gocayne, J.D.; Kerlavage, A.R.; Dougherty, B.A.;
            Tomb, J.F.; Adams, M.D.; Reich, C.I.; Overbeek, R.;
            Kirkness, E.F.; Meinkoth, K.G.; Merrick, J.M.; Glodek, A.;
            Scott, J.L.; Geoghagen, N.S.M.; Weidman, J.F.; Fuhrmann,
            J.L.; Nguyen, D.; Utterback, T.R.; Kelley, J.M.; Peterson,
            J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts,
            K.M.; Hurst, M.A.; Kaine, B.P.; Borodovsky, M.; Klenk,
            H.P.; Fraser, C.M.; Smith, H.O.; Woese, C.R.; Venter, J.C.
            Science (1996) 273:1058-1073
            Complete genome sequence of the methanogenic archaeon,
            Methanococcus jannaschii.
            *cross-references MUID:96337999

```

```
#accession G64403
#status preliminary: nucleic acid sequence not shown;
#translation not shown
#molecule-type DNA
#residues 1-432 #label BUL
#cross-references GB:U57526; GB:L27117; NID:9151508; PDB:gl591518;
#IGR:MJ0831; PDB:gl51089
GENETICS
#map_position F0R750952-752250
#length 432 #molecular-weight 48835 #checksum 5047
SUMMARY
Query Match 76.68; Score 49; DB 2; Length 432;
Best Local Similarity 55.68; Pred. No. 1; Size 01;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
DB 418 F0KRVDSE 426
QY 1 F0KRVSNI 9
```

Search completed: Mon Jun 19 16:06:38 2000  
Job time : 6 secs.

\*\*\*\*\*  
 W E B  
 \*\*\*\*\*  
 (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.  
 Copyright (c) 1993-1998 University of Edinburgh, U.K.  
 Distribution rights by Oxford Molecular Ltd

Mrsrch\_pp protein - protein database search, using Smith-Waterman algorithm  
 Run On: Mon Jun 19 16:05:39 2000; MasPar time 3.40 Seconds  
 Tcellular output not generated; 80,554 Million cell updates/sec

Title: US-09-142-524A-7  
 Description: (1-9) from US9142524A.bef  
 Perfect Score: 64  
 Sequence: : FIKRVSNVI 9  
 Scoring table: PAM 150  
 Gap 15

Searched: 81857 seqs, 30454973 residues

Post-processing: Minimum Match 0%  
 Listing first 45 summaries  
 Database: swiss-prot38  
 l:swissprot

Statistics: Mean 24.343; Variance 25.165; scale 1.967

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Query Match | Length | DB ID        | Description            | Pred. No. |
|------------|-------|-------------|--------|--------------|------------------------|-----------|
| 1          | 64    | 100.0       | 374    | 1 SRP_CRYJA  | SUGI BASIC PROTEIN PRE | 6.81e-04  |
| 2          | 57    | 89.1        | 326    | 1 PEL_EMENI  | PECTATE LYASE PRECURSO | 4.57e-02  |
| 3          | 51    | 79.7        | 720    | 1 YM21_YEAST | HYPOTHETICAL 82.1 KD P | 1.36e-00  |
| 4          | 50    | 78.1        | 129    | 1 CYT1_CYPCA | CYSTATIN PRECURSOR (OV | 2.32e-00  |
| 5          | 50    | 78.1        | 255    | 1 BLAB_BACCE | BETA-LACTAMASE, TYPE I | 2.32e-00  |
| 6          | 49    | 75.6        | 257    | 1 BLA2_BACCE | BETA-LACTAMASE, TYPE I | 3.94e-00  |
| 7          | 49    | 75.6        | 257    | 1 BLA2_BACSP | BETA-LACTAMASE, TYPE I | 3.94e-00  |
| 8          | 49    | 75.6        | 379    | 1 YB21_YEAST | HYPOTHETICAL 45.3 KD P | 3.94e-00  |
| 9          | 49    | 75.6        | 379    | 1 YB21_YEAST | HYPOTHETICAL 45.3 KD P | 3.94e-00  |
| 10         | 49    | 75.6        | 381    | 1 YN76_YEAST | HYPOTHETICAL 45.3 KD P | 3.94e-00  |
| 11         | 49    | 75.6        | 432    | 1 Y831_METJA | HYPOTHETICAL PROTEIN M | 3.94e-00  |
| 12         | 49    | 75.6        | 520    | 1 SYC1_YEAST | CYSTEINYL-TRNA SYNTHET | 3.94e-00  |
| 13         | 49    | 75.6        | 501    | 1 Y066_SCHPO | HYPOTHETICAL 68.8 KD P | 3.94e-00  |
| 14         | 48    | 75.0        | 288    | 1 UB7A_PROST | 4-HYDROXYBENZATE OCTA  | 6.62e-00  |
| 15         | 48    | 75.0        | 4725   | 1 DYHC_DICDI | DYNEIN HEAVY CHAIN, CY | 6.62e-00  |
| 16         | 47    | 73.4        | 509    | 1 KPYK_SCHPO | PYRUVATE KINASE (EC 2. | 1.10e-01  |
| 17         | 47    | 73.4        | 789    | 1 LPB1_DROME | LARVAL SERUM PROTEIN 1 | 1.10e-01  |
| 18         | 47    | 73.4        | 899    | 1 YM31_YEAST | HYPOTHETICAL 103.0 KD  | 1.10e-01  |
| 19         | 46    | 71.9        | 425    | 1 B1CA_SERMA | ADENOSYLMETHIONINE-8-A | 1.82e-01  |
| 20         | 46    | 71.9        | 500    | 1 U24_HCMVA  | HYPOTHETICAL PROTEIN H | 1.82e-01  |
| 21         | 46    | 71.9        | 528    | 1 UDB4_HUMAN | UDP-GLUCURONOSYLTRANSF | 1.82e-01  |
| 22         | 46    | 71.9        | 530    | 1 UDB2_RAT   | UDP-GLUCURONOSYLTRANSF | 1.82e-01  |
| 23         | 46    | 71.9        | 530    | 1 UDB3_RAT   | UDP-GLUCURONOSYLTRANSF | 1.82e-01  |

|    |    |      |      |              |                        |          |
|----|----|------|------|--------------|------------------------|----------|
| 24 | 46 | 71.9 | 530  | 1 UDB5_RAT   | UDP-GLUCURONOSYLTRANSF | 1.82e-01 |
| 25 | 46 | 71.9 | 532  | 1 YRR5_CAEEL | HYPOTHETICAL 59.8 KD P | 1.82e-01 |
| 26 | 46 | 71.9 | 545  | 1 CPT0_LYMST | CYTCHROME P450 1C (EC  | 1.82e-01 |
| 27 | 45 | 71.9 | 585  | 1 SP2E_BACME | STAGE 12 SPOULATION P  | 1.82e-01 |
| 28 | 46 | 71.9 | 810  | 1 MECB_BACSU | NEGATIVE REGULATOR OF  | 1.82e-01 |
| 29 | 46 | 71.9 | 847  | 1 HEX_VIBVU  | BETA-HEXOSAMINIDASE (E | 1.82e-01 |
| 30 | 45 | 71.9 | 857  | 1 CLPB_ECOLL | CLPB PROTEIN (HEAT SHU | 1.82e-01 |
| 31 | 45 | 70.3 | 119  | 1 Y033_ARCFU | SYNAPTOSOMAL ASSOCIATE | 2.98e-01 |
| 32 | 45 | 70.3 | 203  | 1 SN2B_CARAU | SYNAPTOSOMAL ASSOCIATE | 2.98e-01 |
| 33 | 45 | 70.3 | 206  | 1 SN25_HUMAN | SYNAPTOSOMAL ASSOCIATE | 2.98e-01 |
| 34 | 45 | 70.3 | 323  | 1 RTA_FAT    | PROBABLE G PROTEIN-GRU | 2.98e-01 |
| 35 | 45 | 70.3 | 435  | 1 CPXM_BACSU | CYTCHROME P450 1C9 (E  | 2.98e-01 |
| 36 | 45 | 70.3 | 414  | 1 GAG2_DROME | RETROVIRUS-RELATED GAG | 2.98e-01 |
| 37 | 45 | 70.3 | 552  | 1 YHN8_YEAST | HYPOTHETICAL 63.3 KD P | 2.98e-01 |
| 38 | 45 | 70.3 | 574  | 1 DPO2_SCHPO | PROBABLE DNA POLYMERAS | 2.98e-01 |
| 39 | 45 | 70.3 | 586  | 1 KPYK_BACPY | PYRUVATE KINASE (EC 2  | 2.98e-01 |
| 40 | 45 | 70.3 | 587  | 1 KPYK_BACST | PYRUVATE KINASE (EC 2  | 2.98e-01 |
| 41 | 45 | 70.3 | 611  | 1 MEN1_HUMAN | MENIN                  | 2.98e-01 |
| 42 | 45 | 70.3 | 615  | 1 Y083_NPVOP | HYPOTHETICAL 91.1 KD P | 2.98e-01 |
| 43 | 45 | 70.3 | 819  | 1 Y083_METJA | HYPOTHETICAL PROTEIN M | 2.98e-01 |
| 44 | 4  | 70.3 | 1181 | 1 CY_A_SACKL | ADENYATE CYCLASE (EC   | 2.98e-01 |
| 45 | 45 | 70.3 | 1839 | 1            |                        |          |

## ALIGNMENTS

|          |   |                                   |              |
|----------|---|-----------------------------------|--------------|
| RESULT 1 | SBP_CRYJA   | STANDARD:                         | PRT: 374 AA. |
| ID       | AC  | P18632:                           |              |
| DI       | 01-NOV-1990   | (Rel. 16, Created)                |              |
| DI       | 01-NOV-1995   | (Rel. 32, Last sequence update)   |              |
| DI       | 15-DEC-1998   | (Rel. 37, Last annotation update) |              |
| DE       | SUGI: BASIC PROTEIN PRECURSOR (SBP) (MAJOR ALLERGEN CRY J 1) (CRY J 1).   |                                   |              |
| OS       | Cryptomeria japonica (Japanese cedar)   |                                   |              |
| OC       | Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta: euphyllor ytes: Spermatophyta: Coniferopsida: Coniferales: Taxodiaceae: Cryptomeria: [1] |                                   |              |
| RN       | SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.   |                                   |              |
| RP       | TISSUE=POLLEN:  |                                   |              |
| RC       | MEDLINE: 94183234.  |                                   |              |
| RA       | Sone T., Komiyama N., Shimizu K., Kusakabe T., Morikubo K., Kino K.,  |                                   |              |
| RA       | Cloning and sequencing of cDNA coding for Cry J 1, a major allergen of Japanese cedar pollen.   |                                   |              |
| RI       | of Japanese cedar pollen.   |                                   |              |
| RI       | Biochem. Biophys. Res. Commun. 199 419-425(1994)  |                                   |              |
| RN       | [2]   |                                   |              |
| RP       | SEQUENCE FROM N.A.  |                                   |              |
| RC       | TISSUE=POLLEN:  |                                   |              |
| RA       | Narba M., Karose M., Torigoe K., Fikada S., Kurimoto M.,  |                                   |              |
| RL       | Submitted (JUL-1994) to the EMBL/GenBank/DBJ databases.   |                                   |              |
| RN       | [3]   |                                   |              |
| RP       | SEQUENCE OF 22-41.  |                                   |              |
| RC       | TISSUE=POLLEN:  |                                   |              |
| RX       | MEDLINE: 89031257.  |                                   |              |
| RA       | Tanai M., Ando S., Usui M., Kurimoto M., Sakaguchi M., Inouye S.,   |                                   |              |
| RA       | Natural   |                                   |              |
| RI       | "N-terminal amino acid sequence of a major allergen of Japanese cedar pollen (Cry j 1)."  |                                   |              |
| RI       | FEBS Lett. 239:329-332(1988).   |                                   |              |
| RL       | [4]   |                                   |              |
| RN       | CARBOHYDRATES.  |                                   |              |
| RP       | MEDLINE: 95003748.  |                                   |              |
| RA       | Hijikata A., Matsumoto I., Kojima K., Ogawa H.,   |                                   |              |
| RI       | "Antigenicity of the oligosaccharide moiety of the Japanese cedar (Cryptomeria japonica) pollen allergen, Cry j 1."   |                                   |              |
| RL       | Int. Arch. Allergy Immunol. 105:198-202(1994).  |                                   |              |
| CC       | 1- PTM: CONTAINS FUCOSE/XULOSE-CONTAINING N-LINKED OLIGOSACCHARIDES.  |                                   |              |
| CC       | 1- DISEASE: THIS PROTEIN IS THE MAJOR ALLERGEN OF JAPANESE CEDAR.   |                                   |              |
| CC       | 1- POLLEN: THE MOST COMMON POLLEN ALLERGEN IN JAPAN.  |                                   |              |
| CC       | 1- MISCELLANEOUS: THE SEQUENCE OF CRY J 1 FORM A IS SHOWN HERE. FORM B DIFFERS IN SIX POSITIONS.  |                                   |              |
| CC       | 1- SIMILARITY: BELONGS TO THE POLYSACCHARIDE LYASE FAMILY 1.  |                                   |              |



CC -1- FUNCTION: CYSTEINE PROTEINASE INHIBITOR.  
 CC -1- TISSUE SPECIFICITY: DETECTED IN OVARY BUT NOT IN ALL THE OTHER  
 CC EXAMINED TISSUES.  
 CC -1- PIM: THIS PROTEIN IS FURTHER PROCESSED BY AN INTRAMOLECULAR  
 CC CLEAVAGE TO PRODUCE TWO CHAINS LINKED BY A DISULFIDE BRIDGE.  
 CC -1- SIMILARITY: BELONGS TO THE TYPE 2 CYSTATIN FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: L23572; AAB42011.1;  
 CC HSSP: P00388; I490.  
 CC PROSITE: PS00287; CYSTATIN: 1.  
 CC PFAM: PF00031; cystatin: 1.  
 CC TricI protease inhibitor: Signal.  
 CC SIGNAL 1 18  
 CC CHAIN 19 129  
 CC ACT\_SITE 23 23  
 CC REACTIVE SITE (BY SIMILARITY).  
 CC SITE 67 71  
 CC SECONDARY AREA OF CONTACT.  
 CC DISULFID 85 94  
 CC BY SIMILARITY.  
 CC DISULFID 108 128  
 CC BY SIMILARITY.  
 CC SITE 91 92  
 CC CLEAVAGE (PROBABLE).  
 CC SEQUENCE 129 AA: 14236 MW: 732760DF347515BC CRC64;  
 CC  
 CC Query Match 78.1% Score 50; DB 1; Length 129;  
 CC Best Local Similarity 55.6%; Pred. No. 2328-00;  
 CC Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;  
 CC  
 CC 54 FVRKVSQVI 62  
 CC 1:1:1:1:  
 CC 1 FIKRVSNVI 9  
 CC  
 CC RESULT 5  
 CC ID BL21-BACCE STANDARD: PRT: 256 AA.  
 CC AC P14498;  
 CC DT 01-JAN-1990 (Rel. 13, Created)  
 CC DI 03-JAN-1990 (Rel. 13, last sequence update)  
 CC DI 15-JUL-1999 (Rel. 38, last annotation update)  
 CC DE BETA-LACTAMASE, TYPE II; PEPCERSE (EC 3.5.2.6) (PENICILLINASE)  
 CC (CEPHALOSPORINASE).  
 CC OS Bacillus cereus  
 CC OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
 CC OC Bacillus/Staphylococcus group; Bacillus.  
 CC RN 1;  
 CC RP SEQUENCE FROM N.A.  
 CC RX MEDLINE: 88227879;  
 CC RA Lim H.M., Peng J.J., Shaw R.W.;  
 CC R1 "Cloning, nucleotide sequence, and expression of the Bacillus cereus  
 CC 5/5/6 beta-lactamase II structural gene."  
 CC R2 J. Bacteriol. 170:2873-2878(1988).  
 CC RN [2]  
 CC RP MUTAGENESIS  
 CC RX MEDLINE: 9308555;  
 CC RA Lim H.M., Peng J.J.;  
 CC R1 "Mutations affecting the catalytic activity of Bacillus cereus 5/5/6  
 CC beta-lactamase II."  
 CC R2 J. Biol. Chem. 264:11682-11687(1989).  
 CC RN [3]  
 CC RP MUTAGENESIS  
 CC RX MEDLINE: 91264793;  
 CC RA Lim H.M., Iyer R.K., Peng J.J.;  
 CC R1 "Site-directed mutagenesis of dicarboxylic acids near the active site  
 CC of Bacillus cereus 5/5/6 beta-lactamase II."  
 CC R2 Biochem. J. 276:403-404(1991).  
 CC CC -1- FUNCTION: CAN HYDROLYZE CARAPENEM COMPOUNDS.  
 CC -1- CATALYTIC ACTIVITY: A BETA-LACTAM + H(2)O -> A SUBSTITUTED BETA-

CC AMINO ACID.  
 CC -1- COFACTOR: BINDS TWO ZINC IONS PER MOLECULE.  
 CC -1- SIMILARITY: BELONGS TO THE CLASS-B BETA-LACTAMASE FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: M19530; AAA22562.1;  
 CC PIR: A32017; A32017.  
 CC HSSP: P04190; IBM.  
 CC PROSITE: PS00743; BETA-LACTAMASE\_B\_1; 1;  
 CC PROSITE: PS00743; BETA-LACTAMASE\_B\_2; 1;  
 CC PFAM: PF00753; Lactamase\_B\_1;  
 CC Hydrolase; Zinc; Antibiotic resistance; Signal.  
 CC SIGNAL 1 29  
 CC CHAIN 30 256  
 CC METAL 115 115  
 CC BETA-LACTAMASE, TYPE II.  
 CC METAL 117 117  
 CC ZINC 1 (BY SIMILARITY).  
 CC METAL 119 119  
 CC ZINC 2 (BY SIMILARITY).  
 CC METAL 178 178  
 CC ZINC 1 (BY SIMILARITY).  
 CC METAL 197 197  
 CC ZINC 2 (BY SIMILARITY).  
 CC METAL 239 239  
 CC ZINC 2 (BY SIMILARITY).  
 CC MUTAGEN 110 110  
 CC D-N: NO CHANGE IN ACTIVITY.  
 CC MUTAGEN 119 119  
 CC D-N.E: INACTIVATES THE ENZYME.  
 CC MUTAGEN 177 177  
 CC G-D: INACTIVATES THE ENZYME.  
 CC SEQUENCE 256 AA: 28038 MW: 670F52378C35C2D CRC64;  
 CC  
 CC Query Match 78.1% Score 50; DB 1; Length 256;  
 CC Best Local Similarity 66.7%; Pred. No. 2328-00;  
 CC Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;  
 CC  
 CC 104 FKKRVTDVI 112  
 CC 1:1:1:1:  
 CC 1 FIKRVSNVI 9  
 CC  
 CC RESULT 6  
 CC ID BL21-BACCE STANDARD: PRT: 257 AA.  
 CC AC P04190;  
 CC DT 20-MAR-1987 (Rel. 14, Created)  
 CC DI 20-MAR-1987 (Rel. 14, last sequence update)  
 CC DI 15-FEB-2000 (Rel. 39, last annotation update)  
 CC DE BETA-LACTAMASE, TYPE II; PEPCERSE (EC 3.5.2.6) (PENICILLINASE)  
 CC (CEPHALOSPORINASE).  
 CC OS Bacillus cereus  
 CC OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
 CC OC Bacillus/Staphylococcus group; Bacillus.  
 CC RN 1;  
 CC RP SEQUENCE FROM N.A.  
 CC RX MEDLINE: 8600056;  
 CC RA Hussain M., Carlino A., Madocsa M.J., Lampen J.O.;  
 CC R1 "Cloning and sequencing of the metalloprotein beta-lactamase II  
 CC gene of Bacillus cereus 569/H in Escherichia coli."  
 CC R2 J. Bacteriol. 164:223-229(1985).  
 CC RN [2]  
 CC RP SEQUENCE OF 31-183; 187-2. AND 214-257.  
 CC RX STRAIN-569/H;  
 CC RX MEDLINE: 86005446;  
 CC RA Ambler R.P., Daniel M., Fleming J., Hermoso J.M., Pang C., Waley S.G.;  
 CC R1 "The amino acid sequence of the zinc-requiring beta-lactamase II from  
 CC the bacterium Bacillus cereus 569."  
 CC R2 FEBS Lett. 189:207-211(1985).  
 CC RN [3]  
 CC RP X-RAY CRYSTALLOGRAPHY (3.5 ANGSTROMS).  
 CC RX MEDLINE: 9813384;  
 CC RA Sutton B.J., Artymus P.J., Cordero-Borboa A.E., Little G.

RA Phillips D.C., Waley S.G.:  
 "An x-ray-crystallographic study of beta-lactamase II from *Bacillus cereus* at 0.35-nm resolution."  
 Biochem. J. 248:181-188(1987).  
 [4]  
 RX X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).  
 RA MEDLINE: 96067120.  
 RA Carli A., Pares S., Duez E., Galleni M., Duez C., Frere J.-M., Dideberg O.:  
 "The 3-D structure of a zinc metallo-beta-lactamase from *Bacillus cereus* reveals a new type of protein fold."  
 EMBO J. 14:4914-4921(1995).  
 [5]  
 RX X-RAY CRYSTALLOGRAPHY (1.85 ANGSTROMS).  
 RA MEDLINE: 96475525.  
 RA Carli A., Duez E., Galleni M., Frere J.-M., Dideberg O.:  
 "1.85-A resolution structure of the zinc (II) beta-lactamase from *Bacillus cereus*."  
 Acta Crystallogr. D 54:313-323(1998).  
 [6]  
 RX X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).  
 RA MEDLINE: 98400945.  
 RA Fabiano S.M., Soti M.K., Wan T., Payne D.J., Bateman J.H., Mitchell T., Sutton B.J.:  
 "Crystal structure of the zinc-dependent beta-lactamase from *Bacillus cereus* at 1.9-A resolution: binuclear active site with features of a mononuclear enzyme."  
 Biochemistry 37:12404-12411(1998).  
 [7]  
 RX X-RAY CRYSTALLOGRAPHY (1.7 ANGSTROMS).  
 RA STRAIN-559/H:  
 Fabiano S.M., Soti M.K., Sutton B.J.:  
 Submitted (SEP-1997) to the PDB data bank.  
 CC -- FUNCTION: CAN HYDROLYZE CARBAPENEM COMPOUNDS.  
 CC -- CATALYTIC ACTIVITY: A BETA-LACTAM + H(2)O -> A SUBSTITUTED BETA-AMINO ACID.  
 CC -- COFACTOR: BINDS TWO ZINC IONS PER MOLECULE. THE ENZYME CAN ALSO FUNCTION WITH ONLY ONE ZINC ION.  
 CC -- SIMILARITY: BELONGS TO THE CLASS-S BETA-LACTAMASE FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC  
 DE EMBL: M11189; AAA22275.1; 30  
 DE PIR: A31006; PNB592  
 DE PDB: 1BMC; 28-APR-96.  
 DE PDB: 1BME; 07-JUL-97.  
 DE PDB: 1BVI; 23-SEP-98.  
 DE PDB: 1BC2; 14-OCT-98.  
 DE PDB: 28C2; 20-APR-99.  
 DE PDB: 3B2C; 20-APR-99.  
 DE PROSITE: PS00743; BETA\_LACTAMASE\_B\_1; 1.  
 DE PRAM: PF00753; lactamase\_B; 1.  
 DE FUNCTION: zinc: Antibiotic resistance. Signal: 3D-structure.  
 KW Hydrolyase; zinc: Antibiotic resistance. Signal: 3D-structure.  
 FT SIGNAL 1 30  
 FT CHAIN 1 30 BETA-LACTAMASE, TYPE II.  
 FT METAL 116 116 ZINC 1, HIGH-AFFINITY.  
 FT METAL 118 118 ZINC 1, HIGH-AFFINITY.  
 FT METAL 120 120 ZINC 2, LOW-AFFINITY.  
 FT METAL 179 179 ZINC 1, HIGH-AFFINITY.  
 FT METAL 198 198 ZINC 2, LOW-AFFINITY.  
 FT METAL 240 240 ZINC 2, LOW-AFFINITY.  
 SQ SEQUENCE 257 AA: 28092 MW: 25858 B EBF7DDA45431 CRC64;  
 Query Match 76.8% Score 49: DB 1: Length 257;  
 Best Local Similarity 66.7% Pred. No. 3.94e-00;  
 Matches 6: Conservative 2: Mismatches 1: Indels 0: Gaps 0:

Matches 6: Conservative 1: Indels 0: Gaps 0:  
 DB 105 FOKRVTDV1 113  
 QY 1 FIKRVSNV1 9  
 RESULT 7  
 ID BLAZ\_BACSP STANDARD: PRT: 257 AA.  
 AC P04250  
 DT 01-MAR-1989 (Rel. 10, Created)  
 DT 01-MAR-1989 (Rel. 10, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE BETA-LACTAMASE, TYPE II PRAM: PS00743 (EC 3.5.2.6) (PENICILLINASE)  
 OS *Bacillus* sp. (strain 170).  
 OC Bacteria; Firmicutes; Bacilli; Clostridium group;  
 OC *Bacillus*/*Staphylococcus* group; *Bacillus*.  
 RN 1:  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 21-48.  
 RX MEDLINE: 86170399.  
 RA Kato C., Kudo T., Watanabe K., Horikoshi K.:  
 "Nucleotide sequence of the beta-lactamase gene of alkalophilic  
*Bacillus* sp. strain 170."  
 J. Gen. Microbiol. 131:331-332(1985).  
 CC -- FUNCTION: CAN HYDROLYZE CARBAPENEM COMPOUNDS.  
 CC -- CATALYTIC ACTIVITY: A BETA-LACTAM + H(2)O -> A SUBSTITUTED BETA-AMINO ACID.  
 CC -- COFACTOR: BINDS TWO ZINC IONS PER MOLECULE.  
 CC -- SIMILARITY: BELONGS TO THE CLASS-S BETA-LACTAMASE FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC  
 DE EMBL: M15350; AAA22275.1; 30  
 DE PIR: A24393; PNB525.  
 DE HSSP: P04190; LBMC.  
 DE PROSITE: PS00743; BETA\_LACTAMASE\_B\_1; 1.  
 DE PRAM: PS00744; BETA\_LACTAMASE\_B\_2; 1.  
 DE FUNCTION: zinc: Antibiotic resistance. Signal:  
 FT Hydrolyase; zinc: Antibiotic resistance. Signal:  
 FT SIGNAL 1 30  
 FT CHAIN 1 30 BETA-LACTAMASE, TYPE II.  
 FT METAL 116 116 ZINC 1, HIGH-AFFINITY.  
 FT METAL 118 118 ZINC 1, HIGH-AFFINITY.  
 FT METAL 120 120 ZINC 2, LOW-AFFINITY.  
 FT METAL 179 179 ZINC 1, HIGH-AFFINITY.  
 FT METAL 198 198 ZINC 2, LOW-AFFINITY.  
 FT METAL 240 240 ZINC 2, LOW-AFFINITY.  
 SQ SEQUENCE 257 AA: 28153 MW: 23685 B EBFDC045A571 CRC64;  
 Query Match 76.8% Score 49: DB 1: Length 257;  
 Best Local Similarity 66.7% Pred. No. 3.94e-00;  
 Matches 6: Conservative 2: Mismatches 1: Indels 0: Gaps 0:  
 DB 105 FOKRVTDV1 113  
 QY 1 FIKRVSNV1 9  
 RESULT 8  
 ID YFG2\_YEAST STANDARD: PRT: 379 AA.  
 AC P43542;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE HYPOTHEICAL 45.3 KD PROTEIN IN THIS 5'REGION.  
 GN YFG02W.  
 OS *Saccharomyces cerevisiae* (Baker's yeast).

\*Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;  
Saccharomycetaceae; Saccharomyces.

[1]  
SEQUENCE FROM N.A.  
MEDLINE: 95400292.  
RA Murakami Y., Naitou M., Hagiwara H., Shibata T., Ozawa M.,  
RA Sasanuma S.-I., Sasanuma M., Tsuchiya Y., Soeda E., Yokoyama K.,  
RA Yamazaki M., Tashiro H., Eki T.,  
R \*Analysis of the nucleotide sequence of chromosome VI from  
R Saccharomyces cerevisiae.  
R Nat. Genet. 10:251-258(1995).  
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).  
CC -1- SIMILARITY: BELONGS TO THE DUP FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC

EMBL: D50612; BAA09179.1; -  
EMBL: D44503; BAA08057.1; -  
DR PFAM: PF00674; DUP; 2.  
KW Hypothetical protein; Transmembrane.  
FT TRANSMEM 43 63 POTENTIAL.  
FT TRANSMEM 70 90 POTENTIAL.  
FT TRANSMEM 233 253 POTENTIAL.  
FT TRANSMEM 255 275 POTENTIAL.  
SQ SEQUENCE 379 AA: 45312 MW: 622256.969890275 CRC64:

Query Match 76.6% Score 49: DB 1: Length 379;  
Best Local Similarity 55.6% Pred. No. 3.94e+00;  
Matches 5: Conservative 4: Mismatches 0: Indels 0: Gaps 0;

DB 229 FLKRISNIF 237  
QY 1 FIKRVSNVI 9

RESULT 9  
ID YB81-YEAST STANDARD PRI: 479 AA.  
AC P38563.  
DT 01-OCT-1994 (Rel. 30, Created)  
DT 01-OCT-1994 (Rel. 30, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE HYPOTHETICAL 45.2 KD PROTEIN IN WALIS 3' REGION AND IN COX14 5' REGION.  
GN (YBR302C OR YBR2121) AND (YML132W OR YML987.03).  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;  
OC Saccharomycetaceae; Saccharomyces.  
RN [1]  
RP SEQUENCE OF 1-47 FROM N.A. (YBR302C).  
RC STRAIN-S288C.  
RA Feldmann H., Mannhaupt G., Schwarze C., Vetter I.;  
RC Submitted (AUG-1994) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE OF 30-379 FROM N.A. (YBR302C).  
RC STRAIN-S288C.  
RA Feldmann M., Potier S., Souciet J.-L.;  
RN Submitted (AUG-1994) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A. (YML132W).  
RA Bowman S., Louis E.J., Barrell B.G., Rajandream M.A., Walsh S.V.;  
RL Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.  
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).  
CC -1- SIMILARITY: BELONGS TO THE DUP FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC

EMBL: Z36271; CAA85267.1; -  
EMBL: Z50178; CAA90551.1; -  
DR PIR: S46187; S46187.  
DR PFAM: PF00674; DUP; 2.  
KW Hypothetical protein; Transmembrane.  
FT TRANSMEM 73 93 POTENTIAL.  
FT TRANSMEM 255 275 POTENTIAL.  
SQ SEQUENCE 379 AA: 45165 MW: 2AAB51B6C304AF61 CRC64:

Query Match 76.6% Score 49: DB 1: Length 379;  
Best Local Similarity 55.6% Pred. No. 3.94e+00;  
Matches 5: Conservative 4: Mismatches 0: Indels 0: Gaps 0;

DB 229 FLKRISNIF 237  
QY 1 FIKRVSNVI 9

RESULT 10  
ID YN76-YEAST STANDARD PRI: 391 AA.  
AC P53822.  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE HYPOTHETICAL 45.3 KD PROTEIN IN THL12 5' REGION.  
GN YNL336W OR NO275.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;  
OC Saccharomycetaceae; Saccharomyces.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Obermaier B., Piravandi E., Rinke M.;  
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.  
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).  
CC -1- SIMILARITY: BELONGS TO THE DUP FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC

EMBL: Z71612; CAA96270.1; -  
DR PFAM: PF00674; DUP; 2.  
KW Hypothetical protein; Transmembrane.  
FT TRANSMEM 43 63 POTENTIAL.  
FT TRANSMEM 73 93 POTENTIAL.  
FT TRANSMEM 232 252 POTENTIAL.  
FT TRANSMEM 255 275 POTENTIAL.  
SQ SEQUENCE 381 AA: 45288 MW: F9AF930CB1935131 CRC64:

Query Match 76.6% Score 49: DB 1: Length 381;  
Best Local Similarity 55.6% Pred. No. 3.94e+00;  
Matches 5: Conservative 4: Mismatches 0: Indels 0: Gaps 0;

DB 229 FLKRISNIF 237  
QY 1 FIKRVSNVI 9

RESULT 11  
ID Y831-METJA STANDARD PRI: 432 AA.  
AC Q58241.  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)

CC 11- SUBUNIT: MONOMER (BY SIMILARITY).

CC 11- SUBCELLULAR LOCATION: CYTOPLASMIC.

CC 11- SIMILARITY: BELONGS TO CLASS-I AMINOACYL-TRNA SYNTHETASE FAMILY

CC STRONG, TO METHIONYL-TRNA SYNTHETASE.

CC

CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL Outstation

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC

CC EMBL: A5001193; AAC65055.1

CC TIGR: TP0091

CC PRINTS: PR05983; TRNASYNTHCS

CC PROSITE: PS00178; AA-TRNA-LIGASE\_I; FALSE\_NEG.

CC PFAM: PF01426; TRNA-SYNTHASE\_1

CC Aminoacyl-TRNA synthetase: Protein biosynthesis: Ligase, ATP-binding

CC SIMILAR 31 41 "HIGH" REGION

CC SIMILAR 301 305 "KMSK" REGION

CC BINDING 304 304 ATP (BY SIMILARITY)

CC SEQUENCE 520 AA: 58403 MW: 60P2704F7AF2C53E CRC64

Query Match 76.6% Score 49: DB 1: Length 520:

Best Local Similarity 44.4% Pred. No. 3.94e+00:

Matches 4: Conservative 4: Mismatches 1: Indels 0: Gaps 0:

DB 355 FVRRVARVV 363

QY 1 FIKRVSNVI 9

RESULT 13

ID YE06\_SCHPO STANDARD: WT: 631 AA

AC O13603

DI 15-JUL-1998 (Rel. 36, Created)

DI 15-JUL-1998 (Rel. 36, Last sequence update)

DI 15-JUL-1998 (Rel. 36, Last annotation update)

DE HYPOTHEICAL 68.8 KD PROTEIN C17H9.06C IN CHROMOSOME 1

GN SPAC17H9.06C

OS Schizosaccharomyces pombe (Fission yeast)

OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetales

CC Schizosaccharomycetaceae; Schizosaccharomycetes

RN

RP SEQUENCE FROM N.A.

RC STRAIN:972

RA Skellern J., Chatterjee M., Patterson P.G., Patandram H.A., Wood V

RJ Submitted (Aug-1997) to the EMBL/GenBank/CCNU databases

CC

CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL Outstation

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC

CC EMBL: Z7597; CAB1215.1

CC

CC Hypothetical protein.

KW DOMAIN 70 73 POLY-SER.

FT DOMAIN 266 269 POLY-SER.

SQ SEQUENCE 601 AA: 68796 MW: 4694F9296B109D01 CRC64:

Query Match 76.6% Score 49: DB 1: Length 601:

Best Local Similarity 66.7% Pred. No. 3.94e+00:

Matches 6: Conservative 2: Mismatches 1: Indels 0: Gaps 0:

DB 21 FVRRXSNV 29

QY 1 FIKRVSNVI 9

CC 11- SUBUNIT: MONOMER (BY SIMILARITY).

CC 11- SUBCELLULAR LOCATION: CYTOPLASMIC.

CC 11- SIMILARITY: BELONGS TO CLASS-I AMINOACYL-TRNA SYNTHETASE FAMILY

CC STRONG, TO METHIONYL-TRNA SYNTHETASE.

CC

CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL Outstation

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC

CC EMBL: U57526; AAR49830.1

CC TIGR: M30831

CC PRINTS: M30831

CC PROSITE: PF01426; TRNA-SYNTHASE\_1

CC Aminoacyl-TRNA synthetase: Protein biosynthesis: Ligase, ATP-binding

CC SIMILAR 31 41 "HIGH" REGION

CC SIMILAR 301 305 "KMSK" REGION

CC BINDING 304 304 ATP (BY SIMILARITY)

CC SEQUENCE 432 AA: 48895 MW: 0FB95D18D2B41F3 CRC64

Query Match 76.6% Score 49: DB 1: Length 432:

Best Local Similarity 55.6% Pred. No. 3.94e+00:

Matches 5: Conservative 4: Mismatches 1: Indels 0: Gaps 0:

DB 418 FLKRVRIIT 426

QY 1 FIKRVSNVI 9

RESULT 12

ID SVC12PEPA STANDARD: WT: 521 AA

AC G81129

DI 15-DEC-1998 (Rel. 37, Created)

DI 15-DEC-1998 (Rel. 37, Last sequence update)

DI 15-DEC-1998 (Rel. 37, Last annotation update)

DE CYSTEINYL-TRNA SYNTHETASE (EC 6.1.1.16) [CYS-EINP--TRNA LIGASE]

GN CYSYS

OS Treponema pallidum

OC Bacteria; Spirochaetales; Spirochaetaceae; Treponema

RN

RP SEQUENCE FROM N.A.

RC STRAIN:KCH025

RA Fraser C.M., Norris S.J., Weinstock G.M., White O., Sutton G.G.,

RA Dodson R., Gwinn M., Hickey E.K., Clayton R., Ketchum K.A.,

RA Sodergren E., Hardam J.M., McLeod M.P., Salzberg S., Peterson J.,

RA Khalik H., Richardson D., Howell J.K., Chidambaram M., Utterback L.,

RA McDonald L., Artlich P., Bowman C., Cotton M.D., Falli C., Garland S.,

RA Hatch B., Horst K., Roberts K., Sandusky M., Sandusky J., Smith H.O.,

RA Venter J.C.

RJ \*Complete genome sequence of Treponema pallidum, the syphilis

RJ spirochete.

RJ

RJ Science 281:175-188(1998)

CC 11- CATALYTIC ACTIVITY: ATP - L-CYSTEINE - TRNA(CYS) - AMP +

CC PYROPHOSPHATE + L-CYSTEINYL-TRNA(CYS).



```

RESULT 14
ID UBIA_PROST STANDARD: PRT: 288 AA.
AC 052166:
DI 15-JUL-1999 (Rel. 38, Created)
DI 15-JUL-1999 (Rel. 38, Last sequence update)
DI 15-JUL-1999 (Rel. 38, Last annotation update)
DE 4-HYDROXYBENZCOATE OCTAPENYLTRANSFERASE (EC 2.5.1.-) (4-HB
DE POLYPHENYLTRANSFERASE).
GN UBIA OR AARE.
OS Providencia stuartii.
OC Bacteria: Proteobacteria: gamma subdivision: Enterobacteriaceae:
OC Providencia.
RN 1:
RP SEQUENCE FROM N.A.
RC STRAIN=PR50.
RX MEDLINE: 98218723.
RA Paradise M.R., Cook G.M., Poole R.K., Rother P.N.:
RT "Mutations in aar5, the ubia homolog of Providencia stuartii, result
RT in high-level aminoglycoside resistance and reduced expression of the
RT chromosomal aminoglycoside 2'-N-acetyltransferase.";
R1 Antimicrob. Agents Chemother. 42:959-962(1998).
CC -- FUNCTION: SYNTHESIS OF 3-OCTAPENYL-4-HYDROXYBENZCOATE.
CC -- CATALYTIC ACTIVITY: 4-HYDROXYBENZCOATE - FARNESYLARNESYLGERANIOL
CC = 3-OCTAPENYL-4-HYDROXYBENZCOATE.
CC -- COFACTOR: REQUIRES MAGNESIUM (BY SIMILARITY).
CC -- PATHWAY: SECOND STEP IN UBIQUINONE BIOSYNTHESIS.
CC -- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN, INNER MEMBRANE
CC (BY SIMILARITY).
CC -- SIMILARITY: BELONGS TO THE UBIA PRENYLTRANSFERASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement. (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch.)
CC -----
DR EMBL: AF036903; AAB8868.1;
DR PROSITE: PSC0943; UBIA: 1.
DR PFAM: PF01040; CUX10.ctab_cyce; 1.
KW Ubiquinone biosynthesis; Transmembrane; Inner membrane;
KW Magnesium.
FT TRANSMEM 18 58 POTENTIAL.
FT TRANSMEM 108 128 POTENTIAL.
FT TRANSMEM 14 16 POTENTIAL.
FT TRANSMEM 163 183 POTENTIAL.
FT TRANSMEM 213 233 POTENTIAL.
FT TRANSMEM 236 258 POTENTIAL.
FT TRANSMEM 268 288 POTENTIAL.
SQ SEQUENCE 288 AA: 32248 MW: 28819704.18FA403 CRC64;

Query Match 75.0% Score 48: DB 1: Length 288;
Best Local Similarity 55.6% Pred. No. 6.62e-00;
Matches 5: Conservative 4: Mismatches 0: Indels 0: Gaps 0;

Db 134 FVARSVHLL 142
QY 1 FVARSVNI 9

```

```

RC STRAIN=AX3.
RX MEDLINE: 93107159.
RA Koonec M.P., Grissom P.M., McIntosh J.R.:
RT "Dynein from Dictyostellium: primary structure comparisons between a
RT cytoplasmic motor enzyme and flagellar dynein.";
RL J. Cell Biol. 119:1597-1604(1992)
RN 1:
RP CHARACTERIZATION.
RX MEDLINE: 95170719.
RA Koonec M.P., Grissom P.M., Lyon M., Pope T., McIntosh J.R.:
RT "Molecular characterization of a cytoplasmic dynein from
RT Dictyostellium.";
RL J. Eukaryot. Microbiol. 41:645-651(1994).
CC -- FUNCTION: DYNEIN HAS ATPASE ACTIVITY. CYTOPLASMIC DYNEIN ACTS AS A
CC MOTOR FOR THE INTRACELLULAR REORGANIZATION OF VESICLES AND
CC ORGANELLES ALONG MICROTUBULES.
CC -- SUBUNIT: CONSISTS OF AT LEAST TWO HEAVY CHAINS AND A NUMBER OF
CC INTERMEDIATE AND LIGHT CHAINS.
CC -- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -- SIMILARITY: BELONGS TO THE DYNEIN HEAVY CHAIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement. (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch.)
CC -----
DR EMBL: Z15124; CAA78827.1;
DR PIR: S28504; S28504.
DR PIR: A44357; A44357.
DR DICTYDB: D001045; DHCA.
KW Motor protein; Microtubules; Dynein; ATP-binding; Coiled coil
FT DOMAIN 864 897
FT DOMAIN 1199 1219 COILED COIL (POTENTIAL).
FT DOMAIN 1338 1367 COILED COIL (POTENTIAL).
FT DOMAIN 1420 1436 COILED COIL (POTENTIAL).
FT DOMAIN 1656 1684 COILED COIL (POTENTIAL).
FT DOMAIN 2068 2096 MICROTUBULE-BINDING (POTENTIAL)
FT DOMAIN 2226 2248 COILED COIL (POTENTIAL).
FT DOMAIN 2437 2457 COILED COIL (POTENTIAL).
FT DOMAIN 3266 3344 COILED COIL (POTENTIAL).
FT DOMAIN 3438 3580 COILED COIL (POTENTIAL).
FT DOMAIN 3849 3976 ATP (POTENTIAL).
FT NP_BIND 1969 1976 ATP (POTENTIAL).
FT NP_BIND 2271 2278 ATP (POTENTIAL).
FT NP_BIND 2669 2676 ATP (POTENTIAL).
FT NP_BIND 3011 3018 ATP (POTENTIAL).
SQ SEQUENCE 4725 AA: 538807 MW: 94019915D579D3A CRC64;

Query Match 75.0% Score 48: DB 1: Length 4725;
Best Local Similarity 55.6% Pred. No. 6.62e-00;
Matches 5: Conservative 4: Mismatches 0: Indels 0: Gaps 0;

Db 599 YMKRVANVL 707
QY 1 FVARSVNI 9

```

Search completed: Mon Jun 19 16:05:43 2000  
Job time : 4 secs.

\*\*\*\*\*  
[WATER] (TM)  
\*\*\*\*\*  
Release 3.1A John F. Collins, Biocomputing Research Unit,  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

Mwarch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Mon Jun 19 16:06:00 2000: Maspar time 8.22 Seconds  
75,905 Million cell updates/sec  
Tubular output not generated.  
Title: >US-09-142-524A-7  
Description: (1-9) from US09:42524A.pcp  
Perfect Score: 64  
Sequence: 1 FIRVSNVI 9

Scoring table: PAM 150  
Gap 15  
Searched: 225878 seqs, 69334122 residue s  
Post-Processing: Minimum Match 0%  
Listing first 45 summaries  
Database: strembl2  
1:sp\_archaea 2:sp\_bacteria 3:sp\_fungi 4:sp\_human  
5:sp\_invertebrate 6:sp\_mammal 7:sp\_nc 8:sp\_orcellule  
9:sp\_phase 10:sp\_plant 11:sp\_rodent 12:sp\_unclassified  
13:sp\_vertebrate 14:sp\_virus  
Statistics: Mean 23.348; Variance 25.366; scale 0.920

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

| SUMMARIES  |       |              |   |
|------------|-------|--------------|---|
| Result No. | Score | Match Length | ID Description Pred. No.                      |
| 1          | 57    | 89.1         | 2 Q92H39 BETA-N-ACETYLGLUCOSAM 9.55e-02       |
| 2          | 56    | 87.5         | 2 Q92GL9 HYPOTHETICAL 41.5 KD P 1.70e-01      |
| 3          | 53    | 82.8         | 2 Q9XDK3 PUTATIVE FLIPPASE 9.03e-01           |
| 4          | 50    | 78.1         | 2 Q9X273 ENDOGLUCANASE 4.50e-00               |
| 5          | 50    | 78.1         | 2 Q06614 HYPOTHETICAL 50.5 KD P 4.50e-00      |
| 6          | 50    | 78.1         | 2 P71947 HYPOTHETICAL 51.1 KD P 4.50e-00      |
| 7          | 49    | 76.5         | 2 Q9NZ70 TRANSCRIPTIONAL REGULA 7.58e-00      |
| 8          | 49    | 76.6         | 2 Q73300 CYTOCHROME C OXIDASE P 7.58e-00      |
| 9          | 49    | 76.6         | 325 5 Q61959 H05B21.2 PROTEIN 7.58e-00        |
| 10         | 49    | 76.6         | 2 Q924E5 PARA PROTEIN 7.58e-00                |
| 11         | 49    | 76.6         | 451 10 Q82690 MAT1 7.58e-00                   |
| 12         | 49    | 76.6         | 868 13 Q73536 PHEROMONE RECEPTOR 7.58e-00     |
| 13         | 48    | 75.0         | 2 Q9X596 NIDD ALDOLASE 1.27e-01               |
| 14         | 48    | 75.0         | 337 10 Q92NU7 POLLEN MAJOR ALLERGEN 1.27e-01  |
| 15         | 48    | 75.0         | 475 2 Q86020 PYRUVATE KINASE (EC 2. 1.27e-01  |
| 16         | 48    | 75.0         | 482 1 Q58773 CARBAMOYL-PHOSPHATE SY 1.27e-01  |
| 17         | 48    | 75.0         | 601 2 Q07176 HYPOTHETICAL 68.6 KD P 1.27e-01  |
| 18         | 48    | 75.0         | 2342 5 Q01677 P260 1.27e-01                   |
| 19         | 47    | 73.4         | 419 10 Q9XG38 HYPOTHETICAL 51.5 KD P 2.10e-01 |
| 20         | 47    | 73.4         | 508 4 Q915W4 MLL SEPTIN-LIKE FUSION 2.10e-01  |

|    |    |      |      |    |        |    |                        |          |
|----|----|------|------|----|--------|----|------------------------|----------|
| 21 | 47 | 73.4 | 718  | 2  | Q1     | S1 | HYPOTHETICAL 84.7 KD P | 2.10e-01 |
| 22 | 47 | 73.4 | 1452 | 10 | Q1     | S1 | RETROVIRUS-RELATED POL | 2.10e-01 |
| 23 | 46 | 71.9 | 229  | 8  | Q37465 |    | CYTOCHROME C OXIDASE P | 3.45e-01 |
| 24 | 46 | 71.9 | 229  | 8  | Q37559 |    | CYTOCHROME C OXIDASE P | 3.45e-01 |
| 25 | 46 | 71.9 | 229  | 8  | Q37597 |    | CYTOCHROME C OXIDASE P | 3.45e-01 |
| 26 | 46 | 71.9 | 229  | 8  | Q37659 |    | CYTOCHROME C OXIDASE P | 3.45e-01 |
| 27 | 46 | 71.9 | 229  | 8  | Q9XK74 |    | CYTOCHROME C OXIDASE P | 3.45e-01 |
| 28 | 46 | 71.9 | 229  | 8  | Q97566 |    | CYTOCHROME C OXIDASE P | 3.45e-01 |
| 29 | 46 | 71.9 | 229  | 8  | Q97560 |    | CYTOCHROME C OXIDASE P | 3.45e-01 |
| 30 | 46 | 71.9 | 229  | 8  | Q97554 |    | CYTOCHROME C OXIDASE P | 3.45e-01 |
| 31 | 46 | 71.9 | 229  | 8  | Q97560 |    | CYTOCHROME C OXIDASE P | 3.45e-01 |
| 32 | 46 | 71.9 | 229  | 8  | Q97560 |    | CYTOCHROME C OXIDASE P | 3.45e-01 |
| 33 | 46 | 71.9 | 229  | 8  | Q97560 |    | CYTOCHROME C OXIDASE P | 3.45e-01 |
| 34 | 46 | 71.9 | 229  | 8  | Q97560 |    | CYTOCHROME C OXIDASE P | 3.45e-01 |
| 35 | 46 | 71.9 | 229  | 8  | Q97560 |    | CYTOCHROME C OXIDASE P | 3.45e-01 |
| 36 | 46 | 71.9 | 229  | 8  | Q97560 |    | CYTOCHROME C OXIDASE P | 3.45e-01 |
| 37 | 46 | 71.9 | 229  | 8  | Q97560 |    | CYTOCHROME C OXIDASE P | 3.45e-01 |
| 38 | 46 | 71.9 | 229  | 8  | Q97560 |    | CYTOCHROME C OXIDASE P | 3.45e-01 |
| 39 | 46 | 71.9 | 229  | 8  | Q97560 |    | CYTOCHROME C OXIDASE P | 3.45e-01 |
| 40 | 46 | 71.9 | 229  | 8  | Q97560 |    | CYTOCHROME C OXIDASE P | 3.45e-01 |
| 41 | 46 | 71.9 | 229  | 8  | Q97560 |    | CYTOCHROME C OXIDASE P | 3.45e-01 |
| 42 | 46 | 71.9 | 229  | 8  | Q97560 |    | CYTOCHROME C OXIDASE P | 3.45e-01 |
| 43 | 46 | 71.9 | 229  | 8  | Q97560 |    | CYTOCHROME C OXIDASE P | 3.45e-01 |
| 44 | 46 | 71.9 | 229  | 8  | Q97560 |    | CYTOCHROME C OXIDASE P | 3.45e-01 |
| 45 | 46 | 71.9 | 229  | 8  | Q97560 |    | CYTOCHROME C OXIDASE P | 3.45e-01 |

ALIGNMENTS

| RESULT 1   |   |   |              |
|--|---|---|--------------|
| ID   | Q92H39  | PRELIMINARY:                            | PRI: 783 AA. |
| DI   | Q92H39  | 01-MAY-1999 (TrEMBLrel. 10, Created)    |              |
| DI   | Q1-MAY-1999   | (TrEMBLrel. 10, Last sequence update)   |              |
| DI   | Q1-NOV-1999   | (TrEMBLrel. 12, Last annotation update) |              |
| DE   | BETA-N-ACETYLGLUCOSAMINIDASE  |   |              |
| GN   | CHI2  |   |              |
| OS   | Pseudomonas sp. S9.   |   |              |
| OC   | Bacteria: Proteobacteria: gamma subdivision: Alteromonadaceae.      |   |              |
| OC   | Pseudomonas   |   |              |
| RN   | [...]   |   |              |
| RP   | SEQUENCE FROM N.A.  |   |              |
| RC   | STRAIN-S9.  |   |              |
| RA   | TECHNICAL/ANALYST S. GOODMAN  |   |              |
| RT   | Cloning, sequence analysis, characterization of genes involved in   |   |              |
| RT   | cellulose degradation of a marine bacterium, Pseudomonas sp. strain |   |              |
| RT   | S9.   |   |              |
| RL   | Microbiology 010-6(1999)  |   |              |
| DR   | EMBL: AF072374 AAC8237.1  |   |              |
| DR   | HSSP: P06855.10BC   |   |              |
| SC   | SEQUENCE 783 AA: 88856 MW: FF97EDF8 CRC32:                          |   |              |
| Query Match 89.1% Score 57: DB 2: Length 783:              |   |   |              |
| Best Local Similarity 77.8% Pred. No. 9.65e-02:            |   |   |              |
| Matches 7: Conservative 1: Mismatches 1: Indels 0: Gaps 0: |   |   |              |
| Db   | 391 FIRVSOII 399  |   |              |
| Qy   | 1 FIRVSNVI 9  |   |              |
| RESULT 2   |   |   |              |
| ID   | Q92GL9  | PRELIMINARY:                            | PRI: 360 AA. |
| DI   | Q92GL9  | 01-MAY-1999 (TrEMBLrel. 10, Created)    |              |
| DI   | Q1-MAY-1999   | (TrEMBLrel. 10, Last sequence update)   |              |
| DI   | Q1-MAY-1999   | (TrEMBLrel. 10, Last annotation update) |              |
| DE   | HYPOTHETICAL 41.5 KD PROTEIN  |   |              |
| OS   | Leptospira borgpetersenii   |   |              |
| OC   | Bacteria: Spirochaetales: Leptospiraceae: Leptospira                |   |              |
| RN   | [...]   |   |              |
| RP   | SEQUENCE FROM N.A.  |   |              |
| RC   | STRAIN-U171   |   |              |

RA KALAMRAHETTI T., BULACH D.M., RAJAKUMAR K., ADLER B.:  
 "Genetic Organization of the Lipopolysaccharide O-Antigen Biosynthetic  
 Locus of *Leptospira borgpetersenii* Serovar Hardjovitois";  
 Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.

DR EMBL: AF078135; A0012950.1;  
 KW Hypothetical protein.  
 SQ SEQUENCE 360 AA: 41544 MW: 52201263 CRC32:

Query Match 87.5% Score 56; DB 2; Length 360;

Best Local Similarity 77.8%; Pred. NO. 1706-01;  
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

DB 128 FIKRVSNV 125

QY 1 FIKRVSNV 9

RESULT 3

ID Q9XKJ3 PRELIMINARY: PRT: 511 AA.

AC Q9XKJ3

DI 01-NOV-1999 (TRENDEL. 12, Created)

DT 01-NOV-1999 (TRENDEL. 12, Last sequence update)

DE 01-NOV-1999 (TRENDEL. 12, Last annotation update)

DE PUTATIVE FLIPASE.

GN WZX.

OS Bacteroides fragilis.

OC Bacteria: Cytophagales; Bacteroidaceae; Bacteroides.

CC [1]

RP STRAIN-NCTC 9343.

RX MEDLINE: 99307214.

RA COMSTOCK L.E., CAYNE M.J., TZIANABOS A.C., PANTOSI A.,

ONDERDONK A.B., KASPER D.L.:

"Analysis of a capsular polysaccharide biosynthesis locus of

Bacteroides fragilis";

Infect. Immun. 67:3525-3532(1999).

DR EMBL: AF048749; A0040712.1;

SQ SEQUENCE 511 AA: 58642 MW: 54196588 CRC32:

Query Match 82.8% Score 53; DB 2; Length 511;

Best Local Similarity 56.7%; Pred. NO. 9036-01;

Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

DB 501 FIKRVSNV 504

QY 1 FIKRVSNV 9

RESULT 4

ID Q9X273 PRELIMINARY: PRT: 417 AA.

AC Q9X273

DI 01-NOV-1999 (TRENDEL. 12, Created)

DT 01-NOV-1999 (TRENDEL. 12, Last sequence update)

DE 01-NOV-1999 (TRENDEL. 12, Last annotation update)

DE ENDOLUCINASE.

GN TM1752.

OS Thermotoga maritima.

OC Bacteria; Thermotogales; Thermotoga.

CC [1]

RP SEQUENCE FROM N.A.

RX MEDLINE: 99287316

RA HART D.H., HICKEY E.K., PETERSON J.D., NELSON M.C., KETCHUM K.A.,

MCDONALD L., UTTERBACK T.R., MALEK J.A., LINHER K.D., GARRETT M.M.,

RA STEWART A.M., COTTON M.D., PRATT M.S., PHILLIPS C.A., RICHARDSON D.,

RA HEIDELBERG J., SUTTON G.G., FLEISCHMANN R.D., WHITE O., SALZBERG S.L.,

RA SMITH H.O., VENTER J.C., FRASER C.M.:

"Evidence for lateral gene transfer between Archaea and bacteria from

genome sequence of *Thermotoga maritima*.";

Nature 393:323-324(1999).

DR [2]

OS SEQUENCE FROM N.A.

RA NELSON K.E., CLAYTON R.A., GILL S.R., GWINN M.L., DODSON R.J.,

RA HART D.H., HICKEY E.K., PETERSON J.D., NELSON M.C., KETCHUM K.A.,  
 MCDONALD L., UTTERBACK T.R., MALEK J.A., LINHER K.D., GARRETT M.M.,  
 RA STEWART A.M., COTTON M.D., PRATT M.S., PHILLIPS C.A., RICHARDSON D.,  
 RA HEIDELBERG J., SUTTON G.G., FLEISCHMANN R.D., WHITE O., SALZBERG S.L.,  
 RA SMITH H.O., VENTER J.C., FRASER C.M.:

Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.

DR EMBL: AE001813; A0036816.1;

SQ SEQUENCE 317 AA: 37363 MW: 3841108 CRC32:

Query Match 78.1% Score 50; DB 2; Length 317;

Best Local Similarity 56.7%; Pred. NO. 4506-00;

Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

DB 72 FIKRVSNV 80

QY 1 FIKRVSNV 9

RESULT 5

ID OC6614 PRELIMINARY: PRT: 473 AA.

AC OC6614

DI 01-JUL-1997 (TRENDEL. 04, Created)

DT 01-JUL-1997 (TRENDEL. 04, Last sequence update)

DE 01-NOV-1998 (TRENDEL. 08, Last annotation update)

DE HYPOTHETICAL 50.5 KD PROTEIN.

GN MICV336.128.

OS Mycobacterium tuberculosis.

OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae.

OC Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium

RP SEQUENCE FROM N.A.

RC STRAIN-H37RV.

RA MURPHY L., HARRIS D.:

Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN-H37RV.

RA BARRELL B.G., RAZANDEAR M.A., FARKHILL J.:

Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN-H37RV.

RX MEDLINE: 95181548.

RA PHILLIPS K.J., PATEL S., FLEISCHMANN R., FLEISCHMANN R.,

RA BALASUBRAMANIAN V., HENY B., PARKER S., GILL S.R., DODSON R.J.,

RA CLAYTON R.A., GILL S.R., GWINN M.L., DODSON R.J.,

RA NELSON K.E., HART D.H., HICKEY E.K., PETERSON J.D., NELSON M.C., KETCHUM K.A.,

RA MCDONALD L., UTTERBACK T.R., MALEK J.A., LINHER K.D., GARRETT M.M.,

RA STEWART A.M., COTTON M.D., PRATT M.S., PHILLIPS C.A., RICHARDSON D.,

RA HEIDELBERG J., SUTTON G.G., FLEISCHMANN R.D., WHITE O., SALZBERG S.L.,

RA SMITH H.O., VENTER J.C., FRASER C.M.:

"Evidence for lateral gene transfer between Archaea and bacteria from

genome sequence of *Thermotoga maritima*.";

Nature 393:323-324(1999).

DR [2]

OS SEQUENCE FROM N.A.

RA NELSON K.E., CLAYTON R.A., GILL S.R., GWINN M.L., DODSON R.J.,

Q1 Actinocyetales: Corynebacterineae: Mycobacteriaceae: Mycobacterium.

RP SEQUENCE FROM N.A.  
 RC STRAIN-H373V:  
 RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: STRONG. TO M.TUBERCULOSIS MTCV336.28.  
 CC EMBL: 280225; CAB02329.1;  
 CC Hypothetical protein.  
 CC SEQUENCE 419 AA: 51099 MW: 50A262E8 CRC32:

Query Match 78.1% Score 50: DB 2: Length 479:

Best Local Similarity 55.7% Pred. No. 4: 9e+00:  
 Matches 6: Conservative 2: Mismatches 1: Indels 0: Gaps 0:

DB 154 FIKRVSNV 162

|||||

1 FIKRVSNV1 9

RESULT 7 PRELIMINARY: PRT: 190 AA.

AC Q9WZTC  
 CC 01-NOV-1999 (TEMBLrel. 12, Created)  
 CC 01-NOV-1999 (TEMBLrel. 12, Last sequence update)  
 CC 01-NOV-1999 (TEMBLrel. 12, Last annotation update)  
 CC TRANSCRIPTIONAL REGULATOR, TET FAMILY.  
 CC TMC023.  
 CC Thermotoga maritima.  
 CC Bacteria: Thermotogales: Thermotoga.  
 CC [1]  
 CC SEQUENCE FROM N.A.  
 CC MEDLINE: 99287316.  
 CC NELSON K.E., CLAYTON R.A., GILL S.R., GWINN M.L., DOBSON R.J.,  
 CC HAFT D.H., HICKEY E.K., PETERSON J.D., NELSON W.C., KETCHUM K.A.,  
 CC MCDONALD L., GIERBACK J.R., MALEK J.A., LINHER K.D., GARRETT M.M.,  
 CC STEWART A.M., COTTON M.D., PRATT M.S., PHILLIPS C.A., RICHARDSON D.,  
 CC HEIDEMBERG J., SUTTON G.G., FLEISCHMANN R.D., WHITE O., SALZBERG S.L.,  
 CC SMITH H.O., VENTER J.C., FRASER C.M.,  
 CC "Evidence for lateral gene transfer between Archaea and bacteria from  
 CC genome sequence of Thermotoga maritima."  
 CC Nature 399:324-329(1999).

DB 126 FIKRVSEL 134  
 |||||  
 1 FIKRVSNV1 9

Query Match 76.6% Score 49: DB 2: Length 190:

Best Local Similarity 65.7% Pred. No. 7: 56e+00:  
 Matches 6: Conservative 2: Mismatches 1: Indels 0: Gaps 0:

RESULT 8 PRELIMINARY: PRT: 227 AA.

AC Q37300  
 CC 01-NOV-1996 (TEMBLrel. 01, Created)  
 CC 01-NOV-1996 (TEMBLrel. 01, Last sequence update)  
 CC 01-NOV-1999 (TEMBLrel. 12, Last annotation update)  
 CC CYTOCHROME C OXIDASE POLYPEPTIDE 11 (BC 1.9.3.1).  
 CC COI1.  
 CC Chorthippus parallelus.  
 CC Mitochondrion.

OC Eukaryota: Metazoa: Arthropoda: Tracheata: Hexapoda: Insecta:  
 OC Pterygota: Neoptera: Orthopteroidea: Orthoptera: Caelifera:  
 OC Acridomorpha: Acridoidea: Acrididae: Gomphocerinae: Chorthippus.

RP SEQUENCE FROM N.A.

RC STRAIN-ESC:

CC MEDLINE: 96244738.

CC SZYMURA J.M., LUNT D.H., HEWITT G.M.

CC "The sequence and structure of the meadow grasshopper (Chorthippus  
 CC parallelus) mitochondrial rRNA. ND2, COI, COII, ATPase and 9 tRNA  
 CC genes."

CC Insect Mol. Biol. 5:127-135(1996).

CC -1- FUNCTION: SUBUNIT I AND II FORM THE FUNCTIONAL CORE OF THE ENZYME  
 CC COMPLEX. ELECTRONS ORIGINATING IN CYTOCHROME C ARE TRANSFERRED VIA  
 CC HEME A AND CU(A) TO THE B-NUCLEAR CENTER FORMED BY HEME A3 AND  
 CC CU(B). SUBUNIT II BINDS CU(A) AND CYTOCHROME C.

CC -1- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C O(2) - 2 H(2)O - 4

CC FERROCYTOCHROME C.

CC -1- COFACTOR: COPPER A AND HEME GROUP.

CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. (MITOCHONDRIAL)

CC EMBL: X95574; CAA64823.1;

CC PROSITE: PS00078; COX2; 1.

CC SFAM: PF0116; COX2; 1.

CC Mitochondrion: Oxidoreductase: Copper: Transmembrane.

FT METAL 163

FT METAL 167

FT METAL 196

FT METAL 200

FT METAL 204

FT METAL 207

FT METAL 207

CC SEQUENCE 227 AA: 25879 MW: 49E2C7D8 CRC32:

Query Match 76.6% Score 49: DB 2: Length 227:

Best Local Similarity 55.8% Pred. No. 7: 59e+00:

Matches 5: Conservative 4: Mismatches 0: Indels 0: Gaps 0:

DB 219 FIKWISKII 227

|||||

1 FIKRVSNV2 9

RESULT 9

ID Q61969 PRELIMINARY: PRT: 325 AA.

AC Q61969

CC 01-NOV-1996 (TEMBLrel. 17, Created)

CC 01-NOV-1996 (TEMBLrel. 17, Last sequence update)

CC 01-NOV-1999 (TEMBLrel. 12, Last annotation update)

CC H05B21.2 PROTEIN.

CC H05B21.2

CC Caenorhabditis elegans.

OC Eukaryota: Metazoa: Nematoda: Secernentea: Rhabditia: Rhabditidae.

OC Rhabditina: Rhabditoidea: Rhabditidae: Peloderinae: Caenorhabditis.

RP SEQUENCE FROM N.A.

RC STRAIN-BRISTOL N2:

CC MEDLINE: 94150718.

CC WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,

CC BONFIELD J., BURTON C., CONNELL M., COPSEY T., COOPER J., CHOLISIN A.,

CC CRAXTON M., DEAR S., DU Z., JUBIN R., FAVELLO A., FULTON L.,

CC GARDNER A., GREEN P., HAWKIN T., HILLIER L., JIER M., JOHNSTON L.,

CC JONES M., KERSHAW J., KRISTEN J., LAISTER N., LATREILLE P.,

CC LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,

CC PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,

CC SHALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,

CC THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,

CC WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.,

CC "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.

CC elegans."

CC Nature 368:32-38(1994).

CC [2]

RP SEQUENCE FROM N.A.

RC STRAIN-BRISTOL N2:

RA MILLER N.  
 RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.

RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BRISTOL N2  
 RA WATERSTON R.  
 RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ Databases.  
 DR EMBL: AF068717; AAC: 7764; 1; 1;  
 DR PFAM: PF01634; 7cm; 5; 1;  
 SQ SEQUENCE 326 AA: 35948 MW: AAE97C26 CRC32:

Query Match: 76.6%; Score 49; DB 5; Length 326;  
 Best Local Similarity 77.8%; Pred. No. 7,58e+00;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

DB 15 FIKRVSNV: 23  
 1:|||||  
 QY 1 FIKRVSNV: 9

RESULT 10  
 ID 0924E5 PRELIMINARY: PRT: 326 AA.  
 AC 0924E5  
 DT 01-MAY-1999 (TrEMBLrel. 10; Created)  
 DT 01-MAY-1999 (TrEMBLrel. 10; Last sequence update)  
 DT 01-MAY-1999 (TrEMBLrel. 10; Last annotation update)  
 DE PARA PROTEIN.  
 GN PARA.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; gamma subdivision: Enterobacteriaceae;  
 CC Escherichia.  
 RN [1]  
 RA SAMPEI G., MIZOBUCHI K.  
 RI "Organization and diversification of plasmid genomes: complete nucleotide sequence of the Colid-P9 genome."  
 RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AB021078; BAA75111; 1; 1;  
 DR PFAM: AB021078; BAA75111; 1; 1;  
 KR Plasmid.  
 KW SEQUENCE 326 AA: 36227 MW: 9246414 CRC32:

Query Match: 76.6%; Score 49; DB 2; Length 326;  
 Best Local Similarity 66.7%; Pred. No. 7,58e+00;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

DB 260 LKRVSNV: 246  
 1:|||||  
 QY 1 FIKRVSNV: 9

RESULT 11  
 ID 082690 PRELIMINARY: PRT: 451 AA.  
 AC 082690  
 DT 01-NOV-1998 (TrEMBLrel. 08; Created)  
 DT 01-NOV-1998 (TrEMBLrel. 08; Last sequence update)  
 DT 01-NOV-1998 (TrEMBLrel. 08; Last annotation update)  
 DE MAT.  
 GN MAT.  
 OS Lepocinclis buetschlii.  
 OC Eukaryota; Eudlenozoa; Euglenida; Euclenales; Lepocinclis.  
 RN [1]  
 RA DOE-SCH N.A., THOMPSON M.C., HALLICK R.B.  
 RI "A matrase-encoding group III twintron is conserved in deeply rooted Euglenoid species: Are group III introns the chicken or the egg?"  
 RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: 759834; CAB16959; 1; 1;  
 SQ SEQUENCE 451 AA: 52368 MW: BE943A2F CRC32:

Query Match: 76.6%; Score 49; DB 10; Length 451;  
 Best Local Similarity 55.6%; Pred. No. 7,58e+00;  
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

DB 325 FLRRVXN: 333  
 1:|||||  
 QY 1 FIKRVSNV: 9

RESULT 12  
 ID 073636 PRELIMINARY: PRT: 868 AA.  
 AC 073636  
 DT 01-AUG-1998 (TrEMBLrel. 07; Created)  
 DT 01-AUG-1998 (TrEMBLrel. 07; Last sequence update)  
 DT 01-NOV-1999 (TrEMBLrel. 12; Last annotation update)  
 DE PHEROMONE RECEPTOR.  
 GN CA02.1.  
 OS Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Actinopterygii.  
 CC Neopterygii; Teleostei; Euteleostei; Acanthopterygii; Perciformes.  
 CC Tetraodontiformes; Tetraodontidae; Tetraodontidae; Fugu  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE 98226788.  
 RA NAITO T., SAITO Y., YAMAMOTO T., NIZAKI Y., TOMURA K., HATAYA M.,  
 RA NAKANISHI S., BRENNER S.  
 RI "Putative pheromone receptors related to the Ca2+-sensing receptor in Fugu."  
 RL Proc. Natl. Acad. Sci. U.S.A. 95:578-581; (1998).  
 DR EMBL: ABC08858; BAA26123; 1; 1;  
 DR PFAM: PF00003; 7cm; 3; 1;  
 DR PFAM: PF01094; ANF receptor; 1;  
 CR PRINTS: PR00248; GFCRMGR.  
 KW Pheromone.  
 SQ SEQUENCE 868 AA: 95531 MW: 94037642 CRC32:

Query Match: 76.6%; Score 49; DB 13; Length 868;  
 Best Local Similarity 75.0%; Pred. No. 7,58e+00;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

DB 262 LKRVSNV: 289  
 1:|||||  
 QY 2 LKRVSNV: 9

RESULT 13  
 ID 09X596 PRELIMINARY: PRT: 251 AA.  
 AC 09X596  
 DT 01-NOV-1999 (TrEMBLrel. 12; Created)  
 DT 01-NOV-1999 (TrEMBLrel. 12; Last sequence update)  
 DT 01-NOV-1999 (TrEMBLrel. 12; Last annotation update)  
 DE NIDULADOLASE.  
 GN NID.  
 OS Rhodococcus sp. 124  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 CC Actinomycetales; Corynebacteriales; Nocardiaceae; Rhodococcus  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=124;  
 RA STEPHANOPOULOS G., SINSKEY A.J.,  
 RA VANIGIMACHI K., LANKENAU E., LESSARD P.A.  
 RI "Isolation and Characterization of Induced Bioconversion Genes from Rhodococcus strain 124."  
 RL Appl. Microbiol. Biotechnol. 0:0-C (1999).  
 DR EMBL: AF121905; AAD25398; 1; 1;  
 SQ SEQUENCE 251 AA: 27374 MW: 22E5ACD5 CRC32:

Query Match: 75.0%; Score 48; DB 2; Length 251;  
 Best Local Similarity 55.6%; Pred. No. 1,27e+01;  
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

DB 74 FTRRVSEV: 82  
 1:|||||  
 QY 1 FIKRVSNV: 9

```

RESULT 14
ID Q92N37 PRELIMINARY: PRI: 367 AA.
AC Q92N37
DT 01-MAY-1999 (IREMBLrel. 10, Created)
DI 01-MAY-1999 (IREMBLrel. 10, Last sequence update)
DI 01-NOV-1999 (IREMBLrel. 12, Last annotation update)
DE POLLEN MAJOR ALLERGEN 1-2.
CS Juniperus ashei (dark white cedar).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Coniferopsida; Coniferales;
OC Taxodiaceae; Juniperus.
RN 1.
RP SEQUENCE FROM N.A.
RA MIDORO-HORIUI T.M., GOLDRUM R.M., KUROSKY A., WOOD T.G.,
RA BROOKS E.G.
RI "Molecular cloning of mountain cedar (Juniperus ashei) pollen major
RI allergen, Jun a 1."
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF106663; AAD03659.1;
DR EMBL: AF106662; AAD03658.1;
DR MENDEL: 36544; Juncas:1088;36544.
DR MENDEL: 36545; Juncas:1088;36545.
SU SEQUENCE 367 AA: 39824 MW: 40208630 CRC32:

Query Match 75.0% Score 48: DB 10: Length 367;
Best Local Similarity 55.6% Pred. No. 1.27e+01;
Matches 5: Conservative 4; Mismatches 0; Indels 0; Gaps 0;

DE 130 FMKXSHV 138
CV : FIKRVSNV 9

RESULT 15
ID Q86020 PRELIMINARY: PRI: 475 AA.
AC Q86020
DT 01-NOV-1998 (IREMBLrel. 08, Created)
DI 01-NOV-1998 (IREMBLrel. 08, Last sequence update)
DI 01-NOV-1999 (IREMBLrel. 12, Last annotation update)
DE PYRUVATE KINASE (EC 2.7.1.40) (PHOSPHOENOLPYRUVATE KINASE)
DE (PHOSPHOENOL TRANSPHOSPHORYLASE).
GN PYK.
OS Zymomonas mobilis.
OC Bacteria; Proteobacteria; alpha subdivision; Zymomonas group;
OC Zymomonas.
RN 1.
RP SEQUENCE FROM N.A.
RA STEINER P., FUSSENBERGER M., BATTLE J.E., SAUER U.
RI "Cloning and expression of the Zymomonas mobilis pyruvate kinase gene
RI in Escherichia coli."
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
CC 1- CATALYTIC ACTIVITY: ATP + PYRUVATE -> ADP + PHOSPHOENOLPYRUVATE.
CP EMBL: AF079586; AAC29104.1;
CP WSSP: P1A178; LPKY.
DR 2FAM: P00224; PK.
DR PRINTS: P01050; PYRUVTKASE.
KW Pyruvate; Transferease.
SU SEQUENCE 475 AA: 51445 MW: 2883CA3B CRC32:

Query Match 75.0% Score 48: DB 2: Length 475;
Best Local Similarity 55.6% Pred. No. 1.27e+01;
Matches 5: Conservative 2; Mismatches 2; Indels 0; Gaps 0;

DE 199 FVORVEDVI 207
CV : FIKRVSNV 9

```

Search completed: Mon Jun 19 16:06:09 2000  
 Job time : 9 secs.

\*\*\*\*\*  
 [M][E][S][S][A][G][E] (TM)  
 \*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit,  
 Copyright (c) 1993-1998 University of Edinburgh, U.K.  
 Distribution rights by Oxford Molecular Ltd

Match\_pp protein - protein database search, using Smith-Waterman algorithm  
 R.M. On: Mon Jun 19 16:08:27 2000; Maspar time 3.97 seconds  
 Tabular output not generated.  
 101,521 Million cell updates/sec

Title: >US-09-142-524A-8  
 Description: (1-17) from US09142524A.pep  
 Perfect Score: 114  
 Sequence: 1 EYLLSARDVLAVVSK 17

Scoring table: PAM 150  
 Gap 15

Searched: 189963 seqs, 23666106 residues

Post-processing: Minimum Match 0%  
 Listing first 45 summaries

Database: a-geneseq35  
 :geneseqp

Statistics: Mean 20.062; Variance 70.029; scale 0.285

Pred. No. is the number of results predicted by chance to have a  
 score greater than or equal to the score of the result being printed,  
 and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Query Match | Length | DB ID  | Description            | Pred. No. |
|------------|-------|-------------|--------|--------|------------------------|-----------|
| 1          | 114   | 100.0       | 93     | R67386 | M. leprae GroES-like p | 2.83e-04  |
| 2          | 103   | 90.4        | 102    | R22364 | GroES structural prote | 4.15e-03  |
| 3          | 79    | 69.3        | 99     | RC5730 | MBP-57 protein         | 1.20e+00  |
| 4          | 70    | 61.4        | 94     | R67388 | GroES-like protein     | 9.21e+00  |
| 5          | 65    | 57.0        | 94     | R67389 | C. perficiens GroES-li | 2.77e-01  |
| 6          | 65    | 57.0        | 97     | R67390 | E. coli GroES-like pro | 2.77e-01  |
| 7          | 65    | 57.0        | 136    | W98289 | H. pylori SHCP 1543 pr | 2.77e-01  |
| 8          | 61    | 53.5        | 118    | R67373 | Heat shock protein HSP | 6.57e-01  |
| 9          | 61    | 53.5        | 118    | R74338 | Helicobacter pylori he | 6.57e-01  |
| 10         | 61    | 53.5        | 118    | W06731 | H. pylori heat shock p | 6.57e-01  |
| 11         | 60    | 52.6        | 1006   | W72256 | Human receptor type ty | 8.13e-01  |
| 12         | 60    | 52.6        | 1006   | W70525 | Human thymus receptor  | 8.13e-01  |
| 13         | 60    | 52.6        | 1021   | W70526 | Human thymus receptor  | 8.13e-01  |
| 14         | 59    | 51.8        | 74     | W43501 | Human secreted protein | 1.01e-02  |
| 15         | 58    | 50.9        | 102    | R13334 | HypA protein           | 1.24e-02  |
| 16         | 58    | 50.9        | 753    | W93495 | N. meningitidis strain | 1.24e-02  |
| 17         | 57    | 50.0        | 861    | W69977 | Staphylococcus aureus  | 1.53e-02  |
| 18         | 57    | 50.0        | 1096   | W01596 | Inositol-1,4,5-triphos | 1.53e-02  |
| 19         | 57    | 50.0        | 1612   | W65088 | R. prowazekii S-layer  | 1.53e-02  |
| 20         | 56    | 48.1        | 609    | W56935 | Oryza sativa starch (b | 1.89e-02  |
| 21         | 55    | 48.2        | 101    | W16679 | Lawsonia intracellular | 2.32e-02  |
| 22         | 55    | 48.2        | 315    | W02585 | G-protein coupled huma | 2.32e-02  |
| 23         | 55    | 48.2        | 315    | R48713 | G-protein coupled huma | 2.32e-02  |

## ALIGNMENTS

RESULT 1  
 ID R67386 standard; Protein: 93 AA.

AC R67386;  
 DT 22-JUN-1995 (first entry)  
 DE M. leprae GroES-like protein  
 KW GroES: immunogen; vaccine; diagnostic; heat shock protein; HSP;  
 KW GroES-like protein; Helicobacter felis.  
 OS Mycobacterium leprae.  
 PN W09426301-A.

PD 24-NOV-1994.

PF 19-MAY-1994; E01625.

PR 19-MAY-1993; EP-401309.

PR 19-NOV-1993; WO-E03259.

PA (INRM) INST NAT SANTE & RECH MEDICALE.

PA (INSP) INST PASTEUR.

PI Ferrero R, Labigne A, Suerbaum S, Thiberge J;

DR WPI: 95-C06797/01.

PT DNA from Helicobacter pylori and Helicobacter felis - used to

PI develop prods. for detection, treatment and prevention of

PI Helicobacter infection

ES Disclosure: Fig. 7B(i-ii); Isapp; English

CC The sequence of the Helicobacter pylori heat shock protein H  
 CC (given in R67373) was compared to that of other GroES-like  
 CC proteins from Legionella pneumophila (R67387), Escherichia coli  
 CC (R67390), Clostridium perfringens (R67389), Mycobacterium leprae  
 CC (R67386) and thermophilic bacterium (R67388), and regions  
 CC of homology were identified.

SQ Sequence 93 AA;

Query Match 100.0%; Score 114; DB 1; Length 93;

Best Local Similarity 100.0%; Pred. No. 2.83e-04;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 77 EYLLSARDVLAVVSK 93

QY 1 EYLLSARDVLAVVSK 17

RESULT 2

ID R22364 standard; Protein: 102 AA.

AC R22364;

DT 29-JUL-1992 (first entry)

DE GroES structural protein;

KW Heat shock protein; groES gene.

OS Streptomyces albus.

PN W09204452-A.

PD 19-MAR-1992.

PI 03-SEP-1991: F03701.  
 PE 10-SEP-1990: FR-01186.  
 PA (INSP ) INST PASTEUR.  
 PJ Mazodier P, Guglielmi G.  
 DE WPI: 92-114358/14.  
 UK NSDB: Q22483.  
 VI Recombinant DNA contg. heat inducible promoter and heterologous  
 PI gene - also vectors, transformed cells and new heat shock  
 PI proteins of Streptococcus albus.  
 PS Disclosure: Fig 5: 50pp; French.  
 QC The sequence is that of the GroES protein which is encoded by the  
 QC structural gene atcES. See also Q22477-Q22486.  
 SL Sequence 102 AA.

Query Match 90.4% Score 103; DB 1; Length 102;  
 Best Local Similarity 76.5% Pred. No. 4,15e-03;  
 Matches 13; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

DB 95 EYLVLARSARVLAIEK 102  
 QY 1 EYLVLARSARVLAIVK 17  
 ||| | | | | | | | | | | | | | | |

RESULT 3  
 AC R05700 standard: protein: 99 AA.

DE 16-AUG-1990 (first entry)  
 DE MPB-57 protein.  
 KW BCG; tuberculosis; MPB-57; ELISA; ds.  
 US Mycobacterium bovis BCG.  
 PN J02052496-A.  
 PI 22-FEB-1990.  
 PI 18-AUG-1988: 205444.  
 PA (AJIN) Aflinoto KK.  
 PE WPI: 90-103125/14.  
 PE N-PSDR: Q23735.  
 PI BCG bacteria derived MPB-57 protein -  
 PI used esp. in diagnosis of tuberculosis.  
 PS Claim 1: Fig 2: 10pp; Japanese.  
 QC Large quantities of MP-57 protein can be derived from a cultured  
 QC transformant and used in diagnosis of tuberculosis by ELISA.  
 SL Sequence 99 AA.

Query Match 89.3% Score 73; DB 1; Length 99;  
 Best Local Similarity 84.6% Pred. No. 1.20e-03;  
 Matches 11; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

DB 93 EYLVLARSARVVG 95  
 QY 1 EYLVLARSARVLA 15  
 ||| | | | | | | | | | |

RESULT 4  
 AC R67380 standard: Protein: 94 AA.

DE 22-JUN-1995 (first entry)  
 DE GroES-like protein.  
 KW Urease; immunogen; vaccine; diagnostic; heat shock protein; HSP;  
 KW GroES-like protein; Helicobacter felis.  
 US Thermophilic bacterium.  
 PN W09426901-A.  
 PI 24-NOV-1994.  
 PI 19-MAY-1994: E01625.  
 PE 19-MAY-1993: EP-401309.  
 PA (INRM ) INST NAT SANTE & RECH MEDICALE.  
 PE (INSP ) INST PASTEUR.  
 PI Ferrero R, Labigne A, Suerbaum S, Thiberge J;  
 PI DNA from Helicobacter pylori and Helicobacter felis - used to  
 PI develop prods. for detection, treatment and prevention of  
 PI Helicobacter infection

PS Disclosure: Fig. 7B(i-ii): 168pp; English.  
 CC The sequence of the Helicobacter pylori heat shock protein H  
 CC (given in R67373) was compared to that of other GroES-like  
 CC proteins from Legionella pneumophila (R67387), Escherichia coli  
 CC (R67390), Clostridium perfringens (R67389), Mycobacterium leprae  
 CC (R67386) and thermophilic bacterium (R67388), and regions  
 CC of homology were identified.  
 SQ Sequence 94 AA.

Query Match 61.4% Score 70; DB 1; Length 94;  
 Best Local Similarity 57.1% Pred. No. 9.21e-00;  
 Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

DB 80 EYLVLARSARVLAIV 93  
 QY 2 EYLVLARSARVLAIV 15  
 ||| | | | | | | | | | | | | | | |

RESULT 5

ID R67389 standard: Protein: 94 AA.  
 AC R67389;  
 DT 22-JUN-1995 (first entry)  
 DE C. perfringens GroES-like protein.  
 KW Urease; immunogen; vaccine; diagnostic; heat shock protein; HSP;  
 KW GroES-like protein; Helicobacter felis.  
 OS Clostridium perfringens.  
 PN W09426901-A.  
 PI 24-NOV-1994.  
 PI 19-MAY-1994: E01625.  
 PE 19-MAY-1993: EP-401309.  
 PR 19-NOV-1993: WO-E03259.  
 PA (INRM ) INST NAT SANTE & RECH MEDICALE.  
 PA (INSP ) INST PASTEUR.  
 PI Ferrero R, Labigne A, Suerbaum S, Thiberge J;  
 PI WPI: 95-006797/01.  
 PI DNA from Helicobacter pylori and Helicobacter felis - used to  
 PI develop prods. for detection, treatment and prevention of  
 PI Helicobacter infection  
 PS Disclosure: Fig. 7B(i-ii): 168pp; English.  
 CC The sequence of the Helicobacter pylori heat shock protein H  
 CC (given in R67373) was compared to that of other GroES-like  
 CC proteins from Legionella pneumophila (R67387), Escherichia coli  
 CC (R67390), Clostridium perfringens (R67389), Mycobacterium leprae  
 CC (R67386) and thermophilic bacterium (R67388), and regions  
 CC of homology were identified.  
 SQ Sequence 94 AA.

Query Match 57.0% Score 65; DB 1; Length 94;  
 Best Local Similarity 60.0% Pred. No. 2.00e-01;  
 Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

DB 79 EYLVLARSARVLAIV 93  
 QY 1 EYLVLARSARVLAIV 15  
 ||| | | | | | | | | | | | | | | |

RESULT 6

ID R67390 standard: Protein: 97 AA.  
 AC R67390;  
 DT 22-JUN-1995 (first entry)  
 DE E. coli GroES-like protein.  
 KW Urease; immunogen; vaccine; diagnostic; heat shock protein; HSP;  
 KW GroES-like protein; Helicobacter felis.  
 OS Escherichia coli.  
 PN W09426901-A.  
 PD 24-NOV-1994.  
 PE 19-MAY-1994: E01625.  
 PR 19-MAY-1993: EP-401309.  
 PR 19-NOV-1993: WO-E03259.  
 PA (INRM ) INST NAT SANTE & RECH MEDICALE.  
 PA (INSP ) INST PASTEUR.  
 PI Ferrero R, Labigne A, Suerbaum S, Thiberge J;  
 PI WPI: 95-006797/01.



PF DNA from Helicobacter pylori and Helicobacter felis - used to  
 PF develop prods. for detection, treatment and prevention of  
 PF Helicobacter infection  
 PS Disclosure: Fig. 7B(1-11): 168pp; English.  
 CC The sequence of the Helicobacter pylori heat shock protein B  
 CC (given in R67373) was compared to that of other Gr-ES-like  
 CC proteins from Legionella pneumophila (R67387), Escherichia coli  
 CC (R67390), Clostridium perfringens (R67389), Mycobacterium leprae  
 CC (R67386) and thermophilic bacterium (R67388), and regions  
 CC of homology were identified.  
 SQ Sequence 97 AA:  
 Query Match 57.0%; Score 65; DB 1; Length 136;  
 Best Local Similarity 50.0%; Pred. No. 2,778e+01;  
 Matches 8; Conservative 5; Mismatches 3; Indels 0; Gaps 0;  
 Db 81 EYVLMSESDIIAIV 95  
 ||||| |||||  
 QY 1 EYVILSARDVLAIV 15  
 ||||| |||||  
 RESULT 7  
 ID W98289 standard; Protein: 136 AA.  
 AC W98289:  
 DT 31-MAR-1999 (first entry)  
 DE H. pylori GHPC 1543 protein.  
 KW GHPC protein; Helicobacter infection; gastroduodenal disease; gastritis;  
 KW peptic ulcer disease.  
 OS Helicobacter pylori.  
 PN WC9823478-A1.  
 PS 08-OCT-1998.  
 PF 01-APR-1998; J06371.  
 PR 29-JUL-1997; US-502615.  
 PK 01-APR-1997; US-833457.  
 PL 24-JUN-1997; US-881227.  
 PA (HUMAN) HUMAN GENOME SCI INC.  
 PE (TNR) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.  
 FI Al. Garza, A. Kieckhefer, H. Miller, C. Oomen RP, Tomb J;  
 DR WPI: 98-542293/45.  
 DR N-PSDB: X14008.  
 CC New isolated Helicobacter polynucleotides - used to develop products  
 CC for the diagnosis, prevention and treatment of Helicobacter  
 CC infections and gastrointestinal diseases  
 CC Claim 8: Page 348-349; 2054pp; English.  
 CC This sequence represents a Helicobacter pylori GHPC protein of the  
 CC invention. The polypeptides can be used for preventing or treating  
 CC Helicobacter infections, and gastroduodenal diseases associated with  
 CC these infections, including acute, chronic, and atrophic gastritis, and  
 CC peptic ulcer diseases, e.g. gastric and duodenal ulcers. They can also be  
 CC used for the production of antibodies. The products can also be used for  
 CC detection and diagnosis.  
 SQ Sequence 136 AA:  
 Query Match 57.0%; Score 65; DB 1; Length 136;  
 Best Local Similarity 50.0%; Pred. No. 2,778e+01;  
 Matches 8; Conservative 5; Mismatches 3; Indels 0; Gaps 0;  
 Db 8 EYVLMSESDIIAIV 23  
 ||||| |||||  
 QY 1 EYVILSARDVLAIV 16  
 ||||| |||||  
 RESULT 8  
 ID R67373 standard; Protein: 118 AA.  
 AC R67373:  
 DT 22-JUN-1995 (first entry)  
 DE Heat shock protein HSPB  
 KW Urease; urea gene; ureB gene; immunogen; vaccine; diagnostic;  
 KW Helicobacter pylori; Heat shock protein; HSP; chaperonin;  
 KW pil2205.  
 OS Helicobacter felis ATCC 49179.  
 PN WO9426901-A.  
 PS 24-NOV-1994.

PF 19-MAY-1994; E01625.  
 PR 18-MAY-1993; EP-401309.  
 PR 19-NOV-1993; WO-E03259.  
 PA (INRM) INST NAT SANTE & RECH MEDICALE.  
 PA (INSP) INST PASTEUR.  
 PI Ferrero R, Labigne A, Suerbaum S, Thiberge J;  
 DR WPI: 95-006797/01.  
 DR N-PSDB: 075321.  
 PT DNA from Helicobacter pylori and Helicobacter felis - used to  
 PT develop prods. for detection, treatment and prevention of  
 PT Helicobacter infection  
 PS Disclosure: Fig. 6: 168pp; English.  
 CC Vaccine compositions include the A and B subunits (given in R67373)  
 CC of H. felis urease encoded by the ureA/ureB gene (475319) region  
 CC of the urease gene cluster of pil2205 (CNOM 1-1355), as well as the  
 CC heat shock proteins HSPA (R67374) and HSPB (R67373) encoded by the  
 CC urease-associated HSP gene cluster region (Q75321) of pil2205 (CNOM  
 CC 1-1356). Recombinant products are expressed in Escherichia coli.  
 SQ Sequence 118 AA:  
 Query Match 53.5%; Score 61; DB 1; Length 118;  
 Best Local Similarity 40.0%; Pred. No. 6,578e+01;  
 Matches 6; Conservative 6; Mismatches 3; Indels 0; Gaps 0;  
 Db 76 EYVWLEEDILGIVG 90  
 ||||| |||||  
 QY 2 EYVILSARDVLAIV 16  
 ||||| |||||  
 RESULT 9  
 ID R74338 standard; Protein: 118 AA.  
 AC R74338:  
 DT 04-NOV-1995 (first entry)  
 DE Helicobacter pylori heat shock protein HSPA.  
 KW Heat shock protein; HSPA; HSPB; urease; UreA; UreB; UreH;  
 KW vaccine; Helicobacter infection.  
 OS Helicobacter pylori.  
 FH Key Location/Qualifiers  
 FI MISC\_difference 92..118  
 FT /note="P: red vaccine component"  
 PN WC9514093-A.  
 PR 26-MAY-1995.  
 PR 19-NOV-1993; E03259.  
 PA (INRM) INST NAT SANTE & RECH MEDICALE.  
 PA (INSP) INST PASTEUR.  
 PI Ferrero R, Labigne A, Suerbaum S;  
 DR WPI: 95-200383/25.  
 DR N-PSDB: Q90181.  
 PT Immunogenic composition against Helicobacter infection - also  
 PT gene fragment(s) and protein(s) from Helicobacter urease gene  
 PT cluster and heat shock proteins(s).  
 PS Claim 17 + 20; Fig 6: 128pp; English.  
 CC The HSPA heat shock protein is a component of a novel immunogenic  
 CC composition capable of inducing protective antibodies against  
 CC Helicobacter infection. The composition may include the HSPB  
 CC protein (R74339), a urease UreA or UreB protein (R74336, R74337)  
 CC or the UreC protein (R74340). The composition is used to prepare  
 CC a vaccine for humans or animals, especially against H. pylori and  
 CC H. felis. Antibodies against the proteins may be used for  
 CC treating Helicobacter infection, and primers/probes to the DNA  
 CC sequence may be used for detection of Helicobacter infection.  
 SQ Sequence 118 AA:  
 Query Match 53.5%; Score 61; DB 1; Length 118;  
 Best Local Similarity 40.0%; Pred. No. 6,578e+01;  
 Matches 6; Conservative 6; Mismatches 3; Indels 0; Gaps 0;  
 Db 76 EYVWLEEDILGIVG 90  
 ||||| |||||  
 QY 2 EYVILSARDVLAIV 16  
 ||||| |||||

**RESULT 10**

|    |  |                       |                     |                     |                                 |                   |
|----|--|-----------------------|---------------------|---------------------|---------------------------------|-------------------|
| DB | W05731 standard; Protein: 118 AA.  | Query Match           | 52.6%               | Score 60:           | DB 1:                           | Length 1006;      |
| AC | W05731   | Best Local Similarity | 51.5%               | Pred. No. 8.13e-01; | Mismatches                      | 2: Indels 0: Gaps |
| DE | H. pylori heat shock protein HspA.   | Matches               | 8: Conservative     |                     |                                 |                   |
| DI | H. pylori heat shock protein HspA.   |                       |                     |                     |                                 |                   |
| KE | Heat shock protein: HspA; HspB; chaperonin; urease; Urea; UreB;  |                       |                     |                     |                                 |                   |
| KW | vaccine; immunogen; antigen; antibody; chronic gastritis; ulcer;   |                       |                     |                     |                                 |                   |
| KX | pili689.   |                       |                     |                     |                                 |                   |
| OS | Helicobacter pylori strain 85P   |                       |                     |                     |                                 |                   |
| FH | Key  |                       |                     |                     |                                 |                   |
| FT | Location/Qualifiers  |                       |                     |                     |                                 |                   |
| FT | domain   |                       |                     |                     |                                 |                   |
| FT | 92..118  |                       |                     |                     |                                 |                   |
| FT | /note= "this region of HspA resembles a metal binding domain and is a preferred fragment for use in compsns. of the invention (claim 3)"   |                       |                     |                     |                                 |                   |
| FN | WO9634624-A1.  |                       |                     |                     |                                 |                   |
| FI | 07-MAY-1996.   |                       |                     |                     |                                 |                   |
| PF | 02-MAY-1996: F01834.   |                       |                     |                     |                                 |                   |
| PP | 19-MAY-1995: US-447277.  |                       |                     |                     |                                 |                   |
| PR | 19-MAY-1995: US-426577.  |                       |                     |                     |                                 |                   |
| PA | (INSE) INST NAT SANTE & RECH MEDICALE.   |                       |                     |                     |                                 |                   |
| PI | (INSE) INST PASTEUR.   |                       |                     |                     |                                 |                   |
| DR | Ferrero RL, Labigne A, Suerbaum S, Thiberge J.   |                       |                     |                     |                                 |                   |
| DP | WPJ: 96-505600/50.   |                       |                     |                     |                                 |                   |
| DR | N-PSDB: 145681.  |                       |                     |                     |                                 |                   |
| PT | New immunogenic compsn. contg. UroC and HspA antigens of Helicobacter... for treatment and prevention of esp. H pylori infection, also new antibodies specific for these antigens.   |                       |                     |                     |                                 |                   |
| PS | Example 3: Page 113: 18pp; English.  |                       |                     |                     |                                 |                   |
| CC | Heat shock proteins HspA (W06731) and HspB (W06732) are products of a Helicobacter py-Cri gene cluster (see also t45681) in plasmid pILL589 (CMCM 1-1356). Recombinant HspA and HspB can be produced in transformed host cells for use (esp. HspA) with Helicobacter urease subunits (see also W06733-30) in novel immunogenic compns. These compsns. can provoke a mucosal response and are useful as vaccines to protect humans or other animals (esp. cats and dogs) against Helicobacter infection, or for treatment of such infections. The C-terminal fragment (see also W06728) of HspA is a preferred component of such compsns. |                       |                     |                     |                                 |                   |
| CC | Sequence 118 AA.   |                       |                     |                     |                                 |                   |
| CC | Query Match  | 53.5%                 | Score 61:           | DB 1:               | Length 118;                     |                   |
| CC | Best Local Similarity  | 40.0%                 | Pred. No. 6.57e-01; | Mismatches          | 6: Mismatches 3: Indels 0: Gaps |                   |
| CC | Matches  | 6: Conservative       |                     |                     |                                 |                   |
| CC | 76 ENKVELEDTIGING 90   |                       |                     |                     |                                 |                   |
| CC | 118..118   |                       |                     |                     |                                 |                   |
| CC | 2 EYLLSARDVLAVV 16   |                       |                     |                     |                                 |                   |
| CC | 118..118   |                       |                     |                     |                                 |                   |
| CC | 76 ENKVELEDTIGING 90   |                       |                     |                     |                                 |                   |
| CC | 118..118   |                       |                     |                     |                                 |                   |
| CC | 2 EYLLSARDVLAVV 16   |                       |                     |                     |                                 |                   |
| CC | 118..118   |                       |                     |                     |                                 |                   |
| CC | 76 ENKVELEDTIGING 90   |                       |                     |                     |                                 |                   |
| CC | 118..118   |                       |                     |                     |                                 |                   |
| CC | 2 EYLLSARDVLAVV 16   |                       |                     |                     |                                 |                   |
| CC | 118..118   |                       |                     |                     |                                 |                   |
| CC | 76 ENKVELEDTIGING 90   |                       |                     |                     |                                 |                   |
| CC | 118..118   |                       |                     |                     |                                 |                   |
| CC | 2 EYLLSARDVLAVV 16   |                       |                     |                     |                                 |                   |
| CC | 118..118   |                       |                     |                     |                                 |                   |
| CC | 76 ENKVELEDTIGING 90   |                       |                     |                     |                                 |                   |
| CC | 118..118   |                       |                     |                     |                                 |                   |
| CC | 2 EYLLSARDVLAVV 16   |                       |                     |                     |                                 |                   |
| CC | 118..118   |                       |                     |                     |                                 |                   |
| CC | 76 ENKVELEDTIGING 90   |                       |                     |                     |                                 |                   |
| CC | 118..118   |                       |                     |                     |                                 |                   |
| CC | 2 EYLLSARDVLAVV 16   |                       |                     |                     |                                 |                   |
| CC | 118..118   |                       |                     |                     |                                 |                   |
| CC | 76 ENKVELEDTIGING 90   |                       |                     |                     |                                 |                   |
| CC | 118..118   |                       |                     |                     |                                 |                   |
| CC | 2 EYLLSARDVLAVV 16   |                       |                     |                     |                                 |                   |
| CC | 118..118   |                       |                     |                     |                                 |                   |
| CC | 76 ENKVELEDTIGING 90   |                       |                     |                     |                                 |                   |
| CC | 118..118   |                       |                     |                     |                                 |                   |
| CC | 2 EYLLSARDVLAVV 16   |                       |                     |                     |                                 |                   |
| CC | 118..118   |                       |                     |                     |                                 |                   |
| CC | 76 ENKVELEDTIGING 90   |                       |                     |                     |                                 |                   |
| CC | 118..118   |                       |                     |                     |                                 |                   |
| CC | 2 EYLLSARDVLAVV 16   |                       |                     |                     |                                 |                   |
| CC | 118..118   |                       |                     |                     |                                 |                   |
| CC | 76 ENKVELEDTIGING 90   |                       |                     |                     |                                 |                   |
| CC | 1  |                       |                     |                     |                                 |                   |

```

RESULT 13
ID W05226 standard; Protein: 1021 AA.
AC W05226;
DE 02-FEB-1999 (first entry)
DD Human thymus receptor tyrosine kinase (TRTK).
KW Thymus receptor tyrosine kinase; TRTK; human; colon carcinoma;
KW testicular carcinoma; pancreatic carcinoma; lung adenocarcinoma;
KW breast carcinoma; hepatocellular carcinoma; cancer;
KW cell proliferation; diagnosis; therapy; prognosis.
DS Homo sapiens.
FH Key Location/Qualifiers
FT Peptide 1..14
FT Protein /label= Sig_peptide
FT Domain /label= Mat_protein
FT Domain 1..1594
FT Domain /label= Extracellular
FT Domain 595..622
FT Domain /label= Transmembrane
FT Domain 623..1021
FT /label= Intracellular
PR W09844111-AL.
PR 08-OCT-1998;
PR 27-MAR-1998; U06321.
PR 28-MAR-1997; US-042856.
PR (HUMA-) HUMAN GENCME SC; INC.
PR Ruben SW, Secret DR;
PR WPI: 98-542702/45.
PR N-PSDB: V33697.
PR New isolated thymus receptor tyrosine kinase - is used to develop
PR products for the diagnosis and treatment of cancers and other
PR diseases involving aberrant cell proliferation
PR Claim 16; Fig 1A-D; 101pp; English.
PR This is the amino acid sequence of 1 of 2 (see also W05225)
PR putative novel human thymus receptor tyrosine kinases (TRTKs), as
PR deduced from the nucleotide sequence of a cDNA clone (see V33697)
PR obtained from a human foetal cell cDNA library. It has a deduced
PR mol.wt. of 111.0 kDa. The novel TRTKs share sequence homology with
PR cdk9. They are thus members of the Eph family of RTKs and
PR potentially involved in diseases resulting in alterations in normal
PR cell proliferation, such as cancer. They are believed to be
PR involved in haematopoietic development. The invention provides
PR TRTK polynucleotides and polypeptides (including the mature protein,
PR soluble polypeptides, the extracellular, transmembrane and
PR intracellular domains, and epitope-bearing portions of such
PR polypeptides). vectors, host cells and recombinant methods of
PR producing these. Also provided are diagnostic methods for detecting
PR disease states associated with the aberrant expression of TRTK and
PR therapeutic methods for treating such disease states, especially
PR cancer. TRTK is believed to be involved in a number of disease
PR states such as cancers, e.g. testicular carcinoma, pancreatic
PR carcinoma, colon carcinoma, lung adenocarcinoma, mammary carcinoma
PR and hepatocellular carcinoma.
PR Sequence 1021 AA:
Query Match 52.6% Score 60; DB 1; Length 1021;
Residual Similarity 62.5%; Pred.No. 8;136+01;
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

DE 606 FLUAAATVLAIV 618
CV 1111 1111
3 YLLLSARDVLAIV 15

RESULT 14
ID W74801 standard; Protein: 74 AA.
AC W74801;
DE 25-JAN-1999 (first entry)
DD Human secreted protein encoded by gene 72 clone HBIA195.
KW Human; secreted protein; testis; tumour; foetal brain tissue;
KW fusion protein; cancer; central nervous system; seizure;

```

```

KW diagnosis: neurodegenerative disease.
OS Homo sapiens.
FH Key Location/Qualifiers
FT Misc-difference 8 label= unknown
FT Misc-difference 23 label= unknown
FT Misc-difference 38 label= unknown
FT Misc-difference 43 label= unknown
FT Misc-difference 44 label= unknown
FT Misc-difference 73 label= unknown
FT W09839448-A2.
PR 11-SEP-1998; J04493.
PR 06-MAR-1998; US-061060.
PR 02-OCT-1997; US-038621.
PR 07-MAR-1997; US-040161.
PR 07-MAR-1997; US-040162.
PR 07-MAR-1997; US-040163.
PR 07-MAR-1997; US-040333.
PR 07-MAR-1997; US-040334.
PR 07-MAR-1997; US-040335.
PR 11-APR-1997; US-040626.
PR 11-APR-1997; US-043311.
PR 11-APR-1997; US-043312.
PR 11-APR-1997; US-043313.
PR 11-APR-1997; US-043314.
PR 11-APR-1997; US-043568.
PR 11-APR-1997; US-043569.
PR 11-APR-1997; US-043576.
PR 11-APR-1997; US-043578.
PR 11-APR-1997; US-043580.
PR 11-APR-1997; US-043659.
PR 11-APR-1997; US-043670.
PR 11-APR-1997; US-043671.
PR 11-APR-1997; US-043672.
PR 11-APR-1997; US-043674.
PR 23-MAY-1997; US-047492.
PR 23-MAY-1997; US-047500.
PR 23-MAY-1997; US-047501.
PR 23-MAY-1997; US-047502.
PR 23-MAY-1997; US-047503.
PR 23-MAY-1997; US-047581.
PR 23-MAY-1997; US-047582.
PR 23-MAY-1997; US-047583.
PR 23-MAY-1997; US-047584.
PR 23-MAY-1997; US-047585.
PR 23-MAY-1997; US-047586.
PR 23-MAY-1997; US-047587.
PR 23-MAY-1997; US-047588.
PR 23-MAY-1997; US-047589.
PR 23-MAY-1997; US-047590.
PR 23-MAY-1997; US-047592.
PR 23-MAY-1997; US-047593.
PR 23-MAY-1997; US-047594.
PR 23-MAY-1997; US-047595.
PR 23-MAY-1997; US-047596.
PR 23-MAY-1997; US-047597.
PR 23-MAY-1997; US-047598.
PR 23-MAY-1997; US-047599.
PR 23-MAY-1997; US-047600.
PR 23-MAY-1997; US-047601.
PR 23-MAY-1997; US-047612.
PR 23-MAY-1997; US-047613.
PR 23-MAY-1997; US-047614.
PR 23-MAY-1997; US-047615.
PR 23-MAY-1997; US-047617.
PR 23-MAY-1997; US-047618.
PR 23-MAY-1997; US-047632.

```



\*\*\*\*\*  
NWREAF  
\*\*\*\*\*  
(TW)

Release 3.1A John F. Collins, BioComputing Research Unit,  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution Rights by Oxford Molecular Ltd

Mfarch\_pp protein - protein database search using Smith-Waterman algorithm  
R.N on: Tue Jun 20 13:35:33 2000: MasPar time 2.33 Seconds  
105.377 Million cell updates/sec  
Tabular output not generated.

Title: >US-09-142-524A-8  
Description: (1-17) from USC9142524A.pep  
Perfect Score: 114  
Sequence: 1 EYILSAROVJAVYSK 17

Scoring table: PAM 150  
Gap 15

Searched: 145341 seqs, 1443740 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: a-issued  
1:5A.COMB 2:5B.COMB 3:6.COMB 4:PCT.COMB 5:backfiles1

Statistics: Mean 18.743: Variance 66.315: scale 0.2H3

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Query Match | Length | ID   | Description                                     | Pred. No. |
|------------|-------------|--------|------|---|-----------|
| 1          | 114         | 100.0  | 93   | US-08-467- Sequence 36, Application US/08457822 | 1.19e-04  |
| 2          | 70          | 61.4   | 94   | US-08-467- Sequence 36, Application US/08457822 | 4.52e-00  |
| 3          | 65          | 57.0   | 94   | US-08-467- Sequence 36, Application US/08457822 | 1.19e-01  |
| 4          | 55          | 57.0   | 97   | US-08-467- Sequence 36, Application US/08457822 | 4.38e-01  |
| 5          | 61          | 58.5   | 118  | US-08-467- Sequence 36, Application US/08457822 | 3.33e-01  |
| 6          | 57          | 50.0   | 344  | US-09-031- Sequence 3, Application US/08457822  | 7.79e-01  |
| 7          | 57          | 50.0   | 1612 | US-08-169- Sequence 2, Application US/08457822  | 1.19e-02  |
| 8          | 55          | 48.2   | 315  | US-08-118- Sequence 34, Application US/08457822 | 1.19e-02  |
| 9          | 55          | 48.2   | 315  | PCT-US91-C Sequence 1, Application US/08457822  | 1.19e-02  |
| 10         | 55          | 48.2   | 612  | US-08-673- Sequence 1, Application US/08457822  | 1.19e-02  |
| 11         | 55          | 48.2   | 973  | US-08-162- Sequence 8, Application US/08457822  | 1.19e-02  |
| 12         | 54          | 47.4   | 327  | PCT-US95-1 Sequence 35, Application US/08457822 | 1.47e-02  |
| 13         | 54          | 47.4   | 327  | US-08-442- Sequence 35, Application US/08457822 | 1.47e-02  |
| 14         | 54          | 47.4   | 572  | US-08-745- Sequence 1, Application US/08457822  | 1.47e-02  |
| 15         | 54          | 47.4   | 997  | US-08-387- Sequence 4, Application US/08457822  | 1.47e-02  |
| 16         | 53          | 46.5   | 789  | US-08-431- Sequence 20, Application US/08457822 | 1.80e-02  |
| 17         | 53          | 46.5   | 789  | US-08-938- Sequence 20, Application US/08457822 | 1.80e-02  |
| 18         | 52          | 45.6   | 245  | US-08-825- Sequence 1, Application US/08457822  | 2.22e-02  |
| 19         | 52          | 45.6   | 367  | US-08-948- Sequence 4, Application US/08457822  | 2.22e-02  |
| 20         | 52          | 45.6   | 367  | US-08-075- Sequence 4, Application US/08457822  | 2.22e-02  |
| 21         | 52          | 45.6   | 367  | US-08-075- Sequence 2, Application US/08457822  | 2.22e-02  |
| 22         | 52          | 45.6   | 367  | US-08-948- Sequence 2, Application US/08457822  | 2.22e-02  |
| 23         | 52          | 45.6   | 367  | PCT-US91-C Sequence 2, Application US/08457822  | 2.22e-02  |

|    |    |      |      |   |   |
|----|----|------|------|---|---|
| 24 | 52 | 45.6 | 375  | 1 | US-08-464- Sequence 22, Application US/08457822 |
| 25 | 52 | 45.6 | 707  | 4 | PCT-US95-1 Sequence 2, Application US/08457822  |
| 26 | 52 | 45.6 | 707  | 2 | US-08-663- Sequence 2, Application US/08457822  |
| 27 | 52 | 45.6 | 797  | 2 | US-08-323- Sequence 1, Application US/08457822  |
| 28 | 52 | 45.6 | 1012 | 3 | US-09-031- Sequence 4, Application US/08457822  |
| 29 | 52 | 45.6 | 1012 | 3 | US-09-031- Sequence 14, Application US/08457822 |
| 30 | 52 | 45.6 | 1012 | 1 | US-08-219- Sequence 19, Application US/08457822 |
| 31 | 52 | 45.6 | 1012 | 1 | US-08-219- Sequence 3, Application US/08457822  |
| 32 | 52 | 45.6 | 1012 | 3 | US-09-031- Sequence 6, Application US/08457822  |
| 33 | 52 | 45.6 | 1012 | 3 | US-09-031- Sequence 2, Application US/08457822  |
| 34 | 52 | 45.6 | 1012 | 3 | US-09-031- Sequence 9, Application US/08457822  |
| 35 | 52 | 45.6 | 1012 | 3 | US-09-031- Sequence 10, Application US/08457822 |
| 36 | 52 | 45.6 | 1012 | 3 | US-09-031- Sequence 12, Application US/08457822 |
| 37 | 52 | 45.6 | 1012 | 3 | US-09-031- Sequence 14, Application US/08457822 |
| 38 | 52 | 45.6 | 1012 | 3 | US-09-031- Sequence 8, Application US/08457822  |
| 39 | 52 | 45.6 | 1012 | 3 | US-09-031- Sequence 7, Application US/08457822  |
| 40 | 52 | 45.6 | 1012 | 3 | US-08-219- Sequence 1, Application US/08457822  |
| 41 | 52 | 45.6 | 1012 | 2 | US-08-219- Sequence 34, Application US/08457822 |
| 42 | 52 | 45.6 | 1012 | 2 | US-08-219- Sequence 4, Application US/08457822  |
| 43 | 52 | 45.6 | 1012 | 2 | US-08-219- Sequence 4, Application US/08457822  |
| 44 | 52 | 45.6 | 1895 | 2 | US-08-619- Sequence 4, Application US/08457822  |
| 45 | 52 | 45.6 | 1895 | 2 | US-08-619- Sequence 4, Application US/08457822  |

ALIGNMENTS

RESULT 1

ID US-08-467-822-15 STANDARD: PRI: 93 AA.

XX XXXXX

DE Sequence 36, Application US/08457822

XX Sequence 36, Application US/08457822

CC Patent No. 5843460

CC GENERAL INFORMATION:

CC APPLICANT: Sabigne, Ann

CC APPLICANT: Sauerbaum, S. Indstried

CC APPLICANT: Ferrero, Richard L.

CC APPLICANT: Thibierge, Jean-Michele.

CC TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS AGAINST

CC TITLE OF INVENTION: HELICOBACTER INFECTION POLYPEPTIDES FOR USE IN THE

CC TITLE OF INVENTION: COMPOSITIONS, AND NUCLEIC ACID SEQUENCES ENCODING SAID

CC NUMBER OF SEQUENCES: 44

CC CORRESPONDENCE ADDRESS:

CC ADDRESSEE: Pharmacia, Rue de la Vierge, 11, 1300 Lausanne, Switzerland

CC ALLOCESSEE: Pharmacia, Rue de la Vierge, 11, 1300 Lausanne, Switzerland

CC CITY: Washington

CC STATE: D.C.

CC COUNTRY: USA

CC ZIP: 20005-9315

CC COMPUTER READABLE FORM:

CC MEDIUM TYPE: Floppy disk

CC COMPUTER: IBM PC compatible

CC OPERATING SYSTEM: PC-DOS/MS-DOS

CC SOFTWARE: Patent in Release #1.0. Version #1.50

CC CURRENT APPLICATION DATA:

CC APPLICATION NUMBER: US/08/467,822

CC FILING DATE: 06-JUN-1995

CC CLASSIFICATION: 435

CC PRIOR APPLICATION DATA:

CC APPLICATION NUMBER: US 08/447,177

CC FILING DATE: 19-MAY-1995

CC CLASSIFICATION: 435

CC PRIOR APPLICATION DATA:

CC APPLICATION NUMBER: US 08/432,697

CC FILING DATE: 02-MAY-1995

CC CLASSIFICATION: 435

CC ATTORNEY/AGENT INFORMATION:



CC TELEPHONE: (202) 408-4000  
 CC TELEFAX: (202) 408-4400  
 CC INFORMATION FOR SEQ ID NO: 39:  
 CC SEQUENCE CHARACTERISTICS:  
 CC LENGTH: 94 amino acids  
 CC TYPE: amino acid  
 CC STRANDEDNESS: single  
 CC TOPOLOGY: linear  
 CC MOLECULE TYPE: protein  
 CC SEQUENCE 94 AA: 10308 MW: 46636 CN:  
 Query Match 57.0% Score 65: DB 2: Length 94:  
 Best Local Similarity 50.0% Pred. No. 1.38e+01:  
 Matches 9: Conservative 2: Mismatches 4: Indels 0: Gaps 0:  
 Db 79 FEYTLRQDDILAIIV 93  
 QY 1 EBYLLSARDVLAV 15  
 II I I I I I I I I I I  
 RESULT 4  
 ID US-09-467-822-40 STANDARD: PRT: 97 AA.  
 XX A2 xxxxxx  
 DE Sequence 40, Application US/08467822  
 XX Sequence 40, Application US/08467822  
 CC Patent No. 5843460  
 CC GENERAL INFORMATION:  
 CC APPLICANT: Labigne, Agnes  
 CC APPLICANT: Sauerbaum, Sebastien  
 CC APPLICANT: Ferrero, Richard L.  
 CC APPLICANT: Thiberge, Jean-Michel  
 CC TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS AGAINST  
 CC TITLE OF INVENTION: HELICOBACTER INFECTION, POLYPEPTIDES FOR USE IN THE  
 CC TITLE OF INVENTION: COMPOSITIONS, AND NUCLEIC ACID SEQUENCES ENCODING SAID  
 CC NUMBER OF SEQUENCES: 44  
 CC CORRESPONDENCE ADDRESS:  
 CC ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &  
 CC ADDRESSEE: Dunner  
 CC STREET: 1300 I Street, N.W.  
 CC CITY: Washington  
 CC STATE: D.C.  
 CC COUNTRY: USA  
 CC ZIP: 20005-3315  
 CC COMPUTER READABLE FORM:  
 CC MEDIUM TYPE: Floppy disk  
 CC OPERATING SYSTEM: PC DOS/MS-DOS  
 CC SOFTWARE: Patent Release #1.0, Version #1.30  
 CC CURRENT APPLICATION DATA:  
 CC APPLICATION NUMBER: US/08/467,822  
 CC FILING DATE: 06-JUN-1995  
 CC CLASSIFICATION: 435  
 CC PRIOR APPLICATION NUMBER:  
 CC APPLICATION NUMBER: US 08/447,177  
 CC FILING DATE: 19-MAY-1995  
 CC CLASSIFICATION: 435  
 CC PRIOR APPLICATION DATA:  
 CC APPLICATION NUMBER: US 08/432,697  
 CC FILING DATE: 02-MAY-1995  
 CC CLASSIFICATION: 435  
 CC ATTORNEY/AGENT INFORMATION:  
 CC NAME: Meyers, Kenneth J.  
 CC REGISTRATION NUMBER: 25,146  
 CC REFERENCE/DOCKET NUMBER: 03495-0137-02000  
 CC TELECOMMUNICATION INFORMATION:  
 CC TELEPHONE: (202) 408-4000  
 CC TELEFAX: (202) 408-4400

CC INFORMATION FOR SEQ ID NO: 40:  
 CC SEQUENCE CHARACTERISTICS:  
 CC LENGTH: 97 amino acids  
 CC TYPE: amino acid  
 CC STRANDEDNESS: single  
 CC TOPOLOGY: linear  
 CC MOLECULE TYPE: protein  
 CC SEQUENCE 97 AA: 10387 MW: 48352 CN:  
 Query Match 57.0% Score 65: DB 2: Length 97:  
 Best Local Similarity 50.0% Pred. No. 1.38e+01:  
 Matches 9: Conservative 3: Mismatches 3: Indels 0: Gaps 0:  
 Db 81 EEVLINSESDILAIIV 95  
 QY 1 EBYLLSARDVLAV 15  
 II I I I I I I I I I I  
 RESULT 5  
 ID US-08-467-822-29 STANDARD: PRT: 118 AA.  
 XX AC xxxxxx  
 DE Sequence 29, Application US/08467822  
 XX Sequence 29, Application US/08467822  
 CC Patent No. 5843460  
 CC GENERAL INFORMATION:  
 CC APPLICANT: Labigne, Agnes  
 CC APPLICANT: Sauerbaum, Sebastien  
 CC APPLICANT: Ferrero, Richard L.  
 CC APPLICANT: Thiberge, Jean-Michel  
 CC TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS AGAINST  
 CC TITLE OF INVENTION: HELICOBACTER INFECTION, POLYPEPTIDES FOR USE IN THE  
 CC TITLE OF INVENTION: COMPOSITIONS, AND NUCLEIC ACID SEQUENCES ENCODING SAID  
 CC NUMBER OF SEQUENCES: 44  
 CC CORRESPONDENCE ADDRESS:  
 CC ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &  
 CC ADDRESSEE: Dunner  
 CC STREET: 1300 I Street, N.W.  
 CC CITY: Washington  
 CC STATE: D.C.  
 CC COUNTRY: USA  
 CC ZIP: 20005-3315  
 CC COMPUTER READABLE FORM:  
 CC MEDIUM TYPE: Floppy disk  
 CC OPERATING SYSTEM: PC DOS/MS-DOS  
 CC SOFTWARE: Patent Release #1.0, Version #1.30  
 CC CURRENT APPLICATION DATA:  
 CC APPLICATION NUMBER: US/08/467,822  
 CC FILING DATE: 06-JUN-1995  
 CC CLASSIFICATION: 435  
 CC PRIOR APPLICATION NUMBER:  
 CC APPLICATION NUMBER: US 08/447,177  
 CC FILING DATE: 19-MAY-1995  
 CC CLASSIFICATION: 435  
 CC PRIOR APPLICATION DATA:  
 CC APPLICATION NUMBER: US 08/432,697  
 CC FILING DATE: 02-MAY-1995  
 CC CLASSIFICATION: 435  
 CC ATTORNEY/AGENT INFORMATION:  
 CC NAME: Meyers, Kenneth J.  
 CC REGISTRATION NUMBER: 25,146  
 CC REFERENCE/DOCKET NUMBER: 03495-0137-02000  
 CC TELECOMMUNICATION INFORMATION:  
 CC TELEPHONE: (202) 408-4000  
 CC TELEFAX: (202) 408-4400  
 CC INFORMATION FOR SEQ ID NO: 29:  
 CC SEQUENCE CHARACTERISTICS:







```

CC      TYPE: AMINO ACID
CC      STRANDEDNESS: UNKNOWN
CC      TOPOLOGY: UNKNOWN
SQ      SEQUENCE 612 AA: 68803 MW: 1972484 CN:

Query Match      48.2% Score 55: DB 2: Length 612:
Best Local Similarity 33.3% Pred. No. 1:19e-02:
Matches      5: Conservative      7: Mismatches      3: Indels      0: Gaps      0:

DE      182 FLVIAAIAIAIAIFX 196
DY      3 YLLSARDVLAVVSK 17

RESULT 11
DE      US-08-162-809-8 STANDARD: PRI: 973 AA.
XX      xxxxxx
CC      Sequence 8, Application US/08/62809
CC      Patent No. 5457048
CC      GENERAL INFORMATION:
CC      APPLICANT: Pasquale, Elena B.
CC      TITLE OF INVENTION: NOVEL EPH-RELATED TYROSINE KINASES,
CC      TITLE OF INVENTION: NUCLEOTIDE SEQUENCES, AND METHODS OF USE
CC      NUMBER OF SEQUENCES: 26
CC      CORRESPONDENCE ADDRESS:
CC      ADDRESSEE: CAMPBELL AND FLORES
CC      STREET: 4470 La Jolla Village Drive, Suite 700
CC      CITY: San Diego
CC      STATE: California
CC      COUNTRY: United States of America
CC      ZIP: 92122
CC      COMPUTER READABLE FORM:
CC      MEDIUM TYPE: Floppy disk
CC      COMPUTER: IBM PC compatible
CC      OPERATING SYSTEM: PC-DOS/MS-DOS
CC      SOFTWARE: Patent in Release #1.25
CC      CURRENT APPLICATION DATA:
CC      FILING DATE:
CC      CLASSIFICATION:
CC      ATTORNEY/AGENT INFORMATION:
CC      NAME: White, John P.
CC      REGISTRATION NUMBER: 28,679
CC      REFERENCE/DOCKET NUMBER: 45185-C-PCT/JPW/MSO
CC      TELECOMMUNICATION INFORMATION:
CC      TELEPHONE: (212) 278-2400
CC      TELEFAX: (212) 391-0525
CC      INFORMATION FOR SEQ ID NO: 1:
CC      SEQUENCE CHARACTERISTICS:
CC      LENGTH: 327 amino acids
CC      TYPE: amino acid
CC      TOPOLOGY: linear
CC      MOLECULE TYPE: protein
SQ      SEQUENCE 327 AA: 35122 MW: 954271 CN:

Query Match      47.4% Score 54: DB 4: Length 420:
Best Local Similarity 33.3% Pred. No. 1:1e-02:
Matches      9: Conservative      1: Mismatches      0: Indels      0: Gaps      0:

DE      543 EYLVSARDVLAVVSK 17
DY      2 EYLVSARDVLAVVSK 17

RESULT 13
DE      US-08-420-235B-35 STANDARD: PRI: 327 AA.
XX      xxxxxx
CC      Sequence 35, Application US/0842235B
CC      Patent No. 5801042
CC      GENERAL INFORMATION:
CC      APPLICANT: Chang, Yuan
CC      APPLICANT: Moore, Patrick S.
CC      TITLE OF INVENTION: UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS
CC      TITLE OF INVENTION: SEQUENCES AND USES THEREOF
CC      NUMBER OF SEQUENCES: 47
CC      CORRESPONDENCE ADDRESS:
CC      ADDRESSEE: Cooper & Dunham LLP

```



CC TOPOLOGY: linear  
CC MOLECULE TYPE: protein  
SC SEQUENCE 997 AA: 103083 MW: 4504141 CN:  
Query Match 47.4% Score 54: DB 2: Length 997:  
Best Local Similarity 80.0% Pred. No. 147e-02:  
Matches 8: Conservative 0: Mismatches 2: Indels 0: Gaps 0:  
Q1 818 LILIDARIVLA 827  
Q2 11 11  
Q3 4 LLSARDVLA 13

Search completed: Tue Jun 20 13:55:38 2000  
Job time : 5 secs



CC CLASSIFICATION: 435  
 CC ATTORNEY/AGENT INFORMATION:  
 CC NAME: Meyers, Kenneth J.  
 CC REGISTRATION NUMBER: 25,146  
 CC REFERENCE/DOCKET NUMBER: 03495, 0137-00000  
 CC TELEPHONE: (202) 408-4000  
 CC TELEFAX: (202) 408-4400  
 CC INFORMATION FOR SEQ ID NO: 36:  
 CC LENGTH: 93 amino acids  
 CC TYPE: amino acids  
 CC STRANDEDNESS: single  
 CC TOPOLOGY: linear  
 CC MOLECULE TYPE: protein  
 CC SEQUENCE 93 AA: 10002 MW: 48030 GN.

Query Match 100.0% Score 114 DB 10 Length 93  
 Best Local Similarity 100.0% Pident. No. 1 00e-04  
 Matches 17 Conservative 0 Mismatches 0 Indels 0 Gaps 0

CC 77 EYLLSARDVLAVSK 93  
 CC 1 EYLLSARDVLAVSK 17

RESULT 2 STANDARD: PFI: 93 AA.  
 CC US-09-142-524A-8  
 CC xxxxxx

Sequence 36, Application US/08432697

Sequence 36, Application US/08432697  
 CC GENERAL INFORMATION:  
 CC APPLICANT: Labigne, Anes  
 CC APPLICANT: Sauerbaum, Sebastien  
 CC APPLICANT: Ferrero, Richard L.  
 CC APPLICANT: Thibierge, Jean-Michel  
 CC TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS AGAINST  
 CC TITLE OF INVENTION: HELICOBACTER INFECTION, POLYPEPTIDES FOR USE IN THE  
 CC TITLE OF INVENTION: COMPOSITIONS, AND NUCLEIC ACID SEQUENCES ENCODING SAID  
 CC TITLE OF INVENTION: POLYPEPTIDES

CC NUMBER OF SEQUENCES: 44  
 CC CORRESPONDENCE ADDRESS:  
 CC ADDRESSEE: Finckh, Henderson, Parabow, Garrett &  
 CC ADDRESSEE: Duner  
 CC STREET: 1400 1 Street, N.W.  
 CC CITY: Washington  
 CC STATE: D.C.  
 CC COUNTRY: USA  
 CC ZIP: 20005-3315  
 CC COMPUTER READABLE FORM:  
 CC MEDIUM TYPE: Floppy disk  
 CC OPERATING SYSTEM: IBM PC Compatible  
 CC SOFTWARE: Patent Release #1.0, Version #1.30  
 CC CURRENT APPLICATION DATA:  
 CC APPLICATION NUMBER: US/08432,697  
 CC FILING DATE: 02-MAY-1995

CC CLASSIFICATION: 424  
 CC ATTORNEY/AGENT INFORMATION:  
 CC NAME: Meyers, Kenneth J.  
 CC REGISTRATION NUMBER: 25,146  
 CC REFERENCE/DOCKET NUMBER: 03495, 0137-00000  
 CC TELEPHONE: (202) 408-4000  
 CC TELEFAX: (202) 408-4400  
 CC INFORMATION FOR SEQ ID NO: 36:  
 CC SEQUENCE CHARACTERISTICS:  
 CC LENGTH: 93 amino acids

CC TYPE: amino acid  
 CC STRANDEDNESS: single  
 CC TOPOLOGY: linear  
 CC MOLECULE TYPE: protein  
 CC SEQUENCE 93 AA: 10002 MW: 48030 GN.  
 CC Query Match 100.0% Score 114 DB 9 Length 93  
 CC Best Local Similarity 100.0% Pident. No. 1 00e-04  
 CC Matches 17 Conservative 0 Mismatches 0 Indels 0 Gaps 0

CC 77 EYLLSARDVLAVSK 93  
 CC 1 EYLLSARDVLAVSK 17

RESULT 3  
 CC US-09-107-532-6969 STANLA PFI: 11 AA

Sequence 6969, Application US/991,7512  
 CC GENERAL INFORMATION:  
 CC APPLICANT: Lynn A. Co. Scott's Stem and David Bush  
 CC TITLE OF INVENTION: NOVEL ACID AND AMINO ACID SEQUENCES RELATING TO  
 CC TITLE OF INVENTION: ENTEROCOCCUS FAECIUM FOR  
 CC TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS  
 CC NUMBER OF SEQUENCES: 7408  
 CC CORRESPONDENCE ADDRESS:  
 CC ADDRESSEE: Genome Therapeutics Corporation  
 CC STREET: 100 Beaver Street  
 CC CITY: Waltham  
 CC STATE: Massachusetts  
 CC COUNTRY: USA  
 CC ZIP: 02154

CC COMPUTER READABLE FORM:  
 CC MEDIUM TYPE: CD-ROM ISO9660

CC COMPUTER:  
 CC OPERATING SYSTEM:  
 CC SOFTWARE:  
 CC CURRENT APPLICATION DATA:  
 CC APPLICATION NUMBER: US/ 991,7512  
 CC FILING DATE:  
 CC PRIOR APPLICATION DATA:  
 CC APPLICATION NUMBER: US/95598  
 CC FILING DATE: May 14, 1995  
 CC PRIOR APPLICATION DATA:  
 CC APPLICATION NUMBER: US/95598  
 CC FILING DATE: May 14, 1995  
 CC ATTORNEY/AGENT INFORMATION:  
 CC NAME: Attiello, Pamela Decker  
 CC REGISTRATION NUMBER: 43,489  
 CC REFERENCE/DOCKET NUMBER: GTC-012  
 CC TELECOMMUNICATION INFORMATION:  
 CC TELEPHONE: (781)893-5007  
 CC TELEFAX: (781)893-8277

CC INFORMATION FOR SEQ ID NO: 6969:  
 CC SEQUENCE CHARACTERISTICS:  
 CC LENGTH: 119 amino acids  
 CC TYPE: amino acid  
 CC TOPOLOGY: linear  
 CC MOLECULE TYPE: protein  
 CC HYPOTHETICAL: YES  
 CC ORIGINAL SOURCE:  
 CC ORGANISM: Enterococcus faecium  
 CC FEATURE:  
 CC NAME/KEY: misc\_feature  
 CC LOCATION: 1...119

CC SEQUENCE 119 AA: 12807 MW: 74418 GN



```

D1 Sequence 4557, Application US/09134000A
D2
D3
D4
D5
D6
D7
D8
D9
D10
D11
D12
D13
D14
D15
D16
D17
D18
D19
D20
D21
D22
D23
D24
D25
D26
D27
D28
D29
D30
D31
D32
D33
D34
D35
D36
D37
D38
D39
D40
D41
D42
D43
D44
D45
D46
D47
D48
D49
D50
D51
D52
D53
D54
D55
D56
D57
D58
D59
D60
D61
D62
D63
D64
D65
D66
D67
D68
D69
D70
D71
D72
D73
D74
D75
D76
D77
D78
D79
D80
D81
D82
D83
D84
D85
D86
D87
D88
D89
D90
D91
D92
D93
D94
D95
D96
D97
D98
D99
D100
D101
D102
D103
D104
D105
D106
D107
D108
D109
D110
D111
D112
D113
D114
D115
D116
D117
D118
D119
D120
D121
D122
D123
D124
D125
D126
D127
D128
D129
D130
D131
D132
D133
D134
D135
D136
D137
D138
D139
D140
D141
D142
D143
D144
D145
D146
D147
D148
D149
D150
D151
D152
D153
D154
D155
D156
D157
D158
D159
D160
D161
D162
D163
D164
D165
D166
D167
D168
D169
D170
D171
D172
D173
D174
D175
D176
D177
D178
D179
D180
D181
D182
D183
D184
D185
D186
D187
D188
D189
D190
D191
D192
D193
D194
D195
D196
D197
D198
D199
D200
D201
D202
D203
D204
D205
D206
D207
D208
D209
D210
D211
D212
D213
D214
D215
D216
D217
D218
D219
D220
D221
D222
D223
D224
D225
D226
D227
D228
D229
D230
D231
D232
D233
D234
D235
D236
D237
D238
D239
D240
D241
D242
D243
D244
D245
D246
D247
D248
D249
D250
D251
D252
D253
D254
D255
D256
D257
D258
D259
D260
D261
D262
D263
D264
D265
D266
D267
D268
D269
D270
D271
D272
D273
D274
D275
D276
D277
D278
D279
D280
D281
D282
D283
D284
D285
D286
D287
D288
D289
D290
D291
D292
D293
D294
D295
D296
D297
D298
D299
D300
D301
D302
D303
D304
D305
D306
D307
D308
D309
D310
D311
D312
D313
D314
D315
D316
D317
D318
D319
D320
D321
D322
D323
D324
D325
D326
D327
D328
D329
D330
D331
D332
D333
D334
D335
D336
D337
D338
D339
D340
D341
D342
D343
D344
D345
D346
D347
D348
D349
D350
D351
D352
D353
D354
D355
D356
D357
D358
D359
D360
D361
D362
D363
D364
D365
D366
D367
D368
D369
D370
D371
D372
D373
D374
D375
D376
D377
D378
D379
D380
D381
D382
D383
D384
D385
D386
D387
D388
D389
D390
D391
D392
D393
D394
D395
D396
D397
D398
D399
D400
D401
D402
D403
D404
D405
D406
D407
D408
D409
D410
D411
D412
D413
D414
D415
D416
D417
D418
D419
D420
D421
D422
D423
D424
D425
D426
D427
D428
D429
D430
D431
D432
D433
D434
D435
D436
D437
D438
D439
D440
D441
D442
D443
D444
D445
D446
D447
D448
D449
D450
D451
D452
D453
D454
D455
D456
D457
D458
D459
D460
D461
D462
D463
D464
D465
D466
D467
D468
D469
D470
D471
D472
D473
D474
D475
D476
D477
D478
D479
D480
D481
D482
D483
D484
D485
D486
D487
D488
D489
D490
D491
D492
D493
D494
D495
D496
D497
D498
D499
D500
D501
D502
D503
D504
D505
D506
D507
D508
D509
D510
D511
D512
D513
D514
D515
D516
D517
D518
D519
D520
D521
D522
D523
D524
D525
D526
D527
D528
D529
D530
D531
D532
D533
D534
D535
D536
D537
D538
D539
D540
D541
D542
D543
D544
D545
D546
D547
D548
D549
D550
D551
D552
D553
D554
D555
D556
D557
D558
D559
D560
D561
D562
D563
D564
D565
D566
D567
D568
D569
D570
D571
D572
D573
D574
D575
D576
D577
D578
D579
D580
D581
D582
D583
D584
D585
D586
D587
D588
D589
D590
D591
D592
D593
D594
D595
D596
D597
D598
D599
D600
D601
D602
D603
D604
D605
D606
D607
D608
D609
D610
D611
D612
D613
D614
D615
D616
D617
D618
D619
D620
D621
D622
D623
D624
D625
D626
D627
D628
D629
D630
D631
D632
D633
D634
D635
D636
D637
D638
D639
D640
D641
D642
D643
D644
D645
D646
D647
D648
D649
D650
D651
D652
D653
D654
D655
D656
D657
D658
D659
D660
D661
D662
D663
D664
D665
D666
D667
D668
D669
D670
D671
D672
D673
D674
D675
D676
D677
D678
D679
D680
D681
D682
D683
D684
D685
D686
D687
D688
D689
D690
D691
D692
D693
D694
D695
D696
D697
D698
D699
D700
D701
D702
D703
D704
D705
D706
D707
D708
D709
D710
D711
D712
D713
D714
D715
D716
D717
D718
D719
D720
D721
D722
D723
D724
D725
D726
D727
D728
D729
D730
D731
D732
D733
D734
D735
D736
D737
D738
D739
D740
D741
D742
D743
D744
D745
D746
D747
D748
D749
D750
D751
D752
D753
D754
D755
D756
D757
D758
D759
D760
D761
D762
D763
D764
D765
D766
D767
D768
D769
D770
D771
D772
D773
D774
D775
D776
D777
D778
D779
D780
D781
D782
D783
D784
D785
D786
D787
D788
D789
D790
D791
D792
D793
D794
D795
D796
D797
D798
D799
D800
D801
D802
D803
D804
D805
D806
D807
D808
D809
D810
D811
D812
D813
D814
D815
D816
D817
D818
D819
D820
D821
D822
D823
D824
D825
D826
D827
D828
D829
D830
D831
D832
D833
D834
D835
D836
D837
D838
D839
D840
D841
D842
D843
D844
D845
D846
D847
D848
D849
D850
D851
D852
D853
D854
D855
D856
D857
D858
D859
D860
D861
D862
D863
D864
D865
D866
D867
D868
D869
D870
D871
D872
D873
D874
D875
D876
D877
D878
D879
D880
D881
D882
D883
D884
D885
D886
D887
D888
D889
D890
D891
D892
D893
D894
D895
D896
D897
D898
D899
D900
D901
D902
D903
D904
D905
D906
D907
D908
D909
D910
D911
D912
D913
D914
D915
D916
D917
D918
D919
D920
D921
D922
D923
D924
D925
D926
D927
D928
D929
D930
D931
D932
D933
D934
D935
D936
D937
D938
D939
D940
D941
D942
D943
D944
D945
D946
D947
D948
D949
D950
D951
D952
D953
D954
D955
D956
D957
D958
D959
D960
D961
D962
D963
D964
D965
D966
D967
D968
D969
D970
D971
D972
D973
D974
D975
D976
D977
D978
D979
D980
D981
D982
D983
D984
D985
D986
D987
D988
D989
D990
D991
D992
D993
D994
D995
D996
D997
D998
D999

```



CC TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS AGAINST  
 CC TITLE OF INVENTION: HELICOBACTER INFECTION, POLYPEPTIDES FOR USE IN THE  
 CC TITLE OF INVENTION: COMPOSITIONS, AND NUCLEIC ACID SEQUENCES ENCODING SAID  
 CC TITLE OF INVENTION: POLYPEPTIDES  
 CC NUMBER OF SEQUENCES: 44  
 CC CORRESPONDENCE ADDRESS:  
 CC ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &  
 CC ADDRESSEE: Dugner  
 CC STREET: 1300 I Street, N.W.  
 CC CITY: Washington  
 CC STATE: D.C.  
 CC COUNTRY: USA  
 CC ZIP: 20005-3315  
 CC COMPUTER READABLE FORM:  
 CC MEDIUM TYPE: Floppy disk  
 CC COMPUTER: IBM PC compatible  
 CC OPERATING SYSTEM: PC-DOS/MS-DOS  
 CC SOFTWARE: PatentIn Release #1.0, Version #1.30  
 CC CURRENT APPLICATION DATA:  
 CC APPLICATION NUMBER: US/08/432,597  
 CC FILING DATE: 02-MAY-1995  
 CC CLASSIFICATION: 424  
 CC ATTORNEY/AGENT INFORMATION:  
 CC NAME: Meyers, Kenneth J.  
 CC REGISTRATION NUMBER: 25,146  
 CC REFERENCE/DOCKET NUMBER: 03435.0137-00000  
 CC TELEPHONE: (202) 408-4000  
 CC TELEFAX: (202) 408-4000  
 CC INFORMATION FOR SEQ ID NO: 39:  
 CC SEQUENCE CHARACTERISTICS:  
 CC LENGTH: 94 amino acids  
 CC TYPE: amino acid  
 CC STRANDEDNESS: single  
 CC TOPOLOGY: linear  
 CC MOLECULE TYPE: protein  
 CC SEQUENCE 94 AA: 10308 MW: 46636 CN:  
 SQ  
 Query Match 57.0% Score 65: DB 9: Length 94:  
 Best Local Similarity 50.0% Pred. No. 7.37e+01:  
 Matches 9: Conservative 2: Mismatches 4: Indels 0: Gaps 0:  
 DB 75 EEVLIMSESDILAIV 93  
 QY 1 EEYLILSARDVLAV 15  
 RESULT 10  
 ID US-08-432-597-40 STANDARD: PRT: 97 AA.  
 XX  
 AC xxxxxx  
 DT  
 DX  
 DE  
 XX  
 XX  
 Sequence 40, Application: US/08432697  
 DE  
 XX  
 Sequence 40, Application US/08432697  
 CC GENERAL INFORMATION:  
 CC APPLICANT: Labigne, Agnes  
 CC APPLICANT: Sauerbaum, Sebastian  
 CC APPLICANT: Ferrero, Richard L.  
 CC APPLICANT: Thibierge, Jean-Michel  
 CC TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS AGAINST  
 CC TITLE OF INVENTION: HELICOBACTER INFECTION, POLYPEPTIDES FOR USE IN THE  
 CC TITLE OF INVENTION: COMPOSITIONS, AND NUCLEIC ACID SEQUENCES ENCODING SAID  
 CC TITLE OF INVENTION: POLYPEPTIDES  
 CC NUMBER OF SEQUENCES: 44  
 CC CORRESPONDENCE ADDRESS:  
 CC ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &  
 CC ADDRESSEE: Dugner  
 CC STREET: 1300 I Street, N.W.  
 CC CITY: Washington  
 CC STATE: D.C.

CC COUNTRY: USA  
 CC ZIP: 20005-3315  
 CC COMPUTER READABLE FORM:  
 CC MEDIUM TYPE: Floppy disk  
 CC COMPUTER: IBM PC compatible  
 CC OPERATING SYSTEM: PC-DOS/MS-DOS  
 CC SOFTWARE: PatentIn Release #1.0, Version #1.30  
 CC CURRENT APPLICATION DATA:  
 CC APPLICATION NUMBER: US/08/432,597  
 CC FILING DATE: 02-MAY-1995  
 CC CLASSIFICATION: 424  
 CC ATTORNEY/AGENT INFORMATION:  
 CC NAME: Meyers, Kenneth J.  
 CC REGISTRATION NUMBER: 25,146  
 CC REFERENCE/DOCKET NUMBER: 03435.0137-00000  
 CC TELEPHONE: (202) 408-4000  
 CC TELEFAX: (202) 408-4000  
 CC INFORMATION FOR SEQ ID NO: 40:  
 CC SEQUENCE CHARACTERISTICS:  
 CC LENGTH: 97 amino acids  
 CC TYPE: amino acid  
 CC STRANDEDNESS: single  
 CC TOPOLOGY: linear  
 CC MOLECULE TYPE: protein  
 CC SEQUENCE 97 AA: 10387 MW: 49352 CN:  
 SQ  
 Query Match 57.0% Score 65: DB 9: Length 97:  
 Best Local Similarity 50.0% Pred. No. 7.37e+01:  
 Matches 9: Conservative 3: Mismatches 3: Indels 0: Gaps 0:  
 DB 81 EEVLIMSESDILAIV 95  
 QY 1 EEYLILSARDVLAV 15  
 RESULT 11  
 ID US-09-472-971-4 STANDARD: PRT: 97 AA.  
 XX  
 AC xxxxxx  
 DT  
 DX  
 DE  
 XX  
 XX  
 Sequence 4, Application: US/09472971  
 CC GENERAL INFORMATION:  
 CC APPLICANT: SOGO, Kazuyo  
 CC APPLICANT: YANAGI, Hideki  
 CC APPLICANT: YURA, Takashi  
 CC TITLE OF INVENTION: TRIGGER FACTOR EXPRESSION PLASMID  
 CC FILE REFERENCE: 1422-409E  
 CC CURRENT APPLICATION NUMBER: US/09/472,971  
 CC CURRENT FILING DATE: 1999-12-29  
 CC EARLIER APPLICATION NUMBER: JP10-372965  
 CC EARLIER FILING DATE: 1998-12-28  
 CC NUMBER OF SEQ ID NOS: 7  
 CC SOFTWARE: PatentIn Ver. 2.1  
 CC SEQ ID NO 4  
 CC LENGTH: 97  
 CC TYPE: PRT  
 CC ORGANISM: Escherichia coli  
 CC SEQUENCE 97 AA: 10387 MW: 4,352 CN:  
 SQ  
 Query Match 57.0% Score 65: DB 20: Length 97:  
 Best Local Similarity 60.0% Pred. No. 7.37e+01:  
 Matches 9: Conservative 3: Mismatches 3: Indels 0: Gaps 0:  
 DB 81 EEVLIMSESDILAIV 95  
 QY 1 EEYLILSARDVLAV 15



CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Clark, Paul T.  
CC REGISTRATION NUMBER: 30,162  
CC REFERENCE/DOCKET NUMBER: 06132/041WO1  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: 617-428-0200  
CC TELEFAX: 617-428-7045  
CC TELEX:

CC INFORMATION FOR SEQ ID NO: 178:

CC SEQUENCE CHARACTERISTICS:

CC LENGTH: 136 amino acids

CC TYPE: amino acid

CC STRANDEDNESS: single

CC TOPOLOGY: linear

CC MOLECULE TYPE: protein

CC FRAGMENT TYPE: internal

CC SEQUENCE 136 AA: 15614 MW: 100717 CN:

Query Match 57.0% Score 65; DB 1; Length 136;

Best Local Similarity 50.3%; Pred. NO. 7.37e-01;

Matches 8; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

CC 8 EBYMLMSREVSFAVG 23

CC 136 1111111111

CC 1 EBYMLMSRDVLAWS 16

RESULT 15

CC US-08-633-457-179 STANDARD: PRI: 136 AA.

CC xxxxxx

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC MOLECULE TYPE: protein  
CC FRAGMENT TYPE: internal  
CC SEQUENCE 136 AA: 15614 MW: 100717 CN:

Query Match 57.0% Score 65; DB 14; Length 136;

Best Local Similarity 50.3%; Pred. NO. 7.37e-01;

Matches 8; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

CC 8 EBYMLMSREVSFAVG 23

CC 136 1111111111

CC 1 EBYMLMSRDVLAWS 16

Search completed: Mon Jun 19 16:12:45 2000

Job time : 21 secs.

Sequence 178. Application US/09833457

Sequence 179. Application US/08833457

GENERAL INFORMATION:

APPLICANT: Harold Kleanthous et al.

TITLE OF INVENTION: Identification of Polynucleotides

TITLE OF INVENTION: Encoding Novel Helicobacter Polypeptides in the Helicobacter

TITLE OF INVENTION: Genome

NUMBER OF SEQUENCES: 370

CORRESPONDENCE ADDRESS:

ADDRESSEE: Clark & Sibling LLP

STREET: 176 Federal Street

CITY: Boston

STATE: MA

COUNTRY: USA

ZIP: 02110

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FASTSEQ for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/833,457

FILING DATE: 01-APR-97

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Clark, Paul T.

REGISTRATION NUMBER: 30,162

REFERENCE/DOCKET NUMBER: 06132/041001

TELECOMMUNICATION INFORMATION:

TELEPHONE: 617-428-0200

TELEFAX: 617-428-7045

TELEX:

INFORMATION FOR SEQ ID NO: 178:

SEQUENCE CHARACTERISTICS:

LENGTH: 136 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear



```

RESULT      2
ENTRY
  TITLE      LK1 T-cell antigen - Mycobacterium leprae
  ORGANISM   Mycobacterium leprae
  DATE       30-Jun-1992
ACCESSIONS  J05556
REFERENCE
  #authors   Mehra, V.; Bloom, B.R.; Bajardi, A.C.; Grisso, C.L.; Sieling, P.A.; Alland, D.; Conville, J.; Fan, X.; Hunter, S.W.; Brennan, P.J.; Rea, T.H.; Modlin, R.L.
  #journal   J. Exp. Med. (1992) 175:275-284
  #title     A major T cell antigen of Mycobacterium leprae is a 10-kD heat-shock cognate protein.
  #cross-references MIMD:92113469
  #accession J05556
  #status    nucleic acid sequence not shown
  #molecule_type DNA
  #residues  1-100 #label MEH
  #cross-references EMBL:X59413
  #comment   This protein is a heat-shock protein.
  #classification #superfamily chaperonin groES
  #summary     #length 100 #molecular-weight 10766 #checksum 7371
  Query Match 100.0% Score 114; DB 2; Length 100;
  Best Local Similarity 100.0%; Pred No. 1.07e-10;
  Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db      84 EBYLILSARDVLAVSK 100
Q:      1 EBYLILSARDVLAVSK 17
      1 |||||.....|||
      2 BWMYRA      #type complete
      3 chaperonin groES - Mycobacterium tuberculosis
      4 LK1 antigen; BCG-a homolog; Cpn10; heat shock protein 10K
      5 (tspl0); immunogenic protein; hcg-a
      6 #normal-name Mycobacterium tuberculosis
      7 #1-Dec-1990 #sequence-revision 31-Dec-1990 #text-change
      8 16-Jul-1999
      9 ACCESSIONS  S03381; S07272; A47292; G70737; A46481; S18040
      10 REFERENCE  S03381
      11 #authors   Baird, P.N.; Hall, L.M.C.; Coates, A.R.M.
      12 #journal   Nucleic Acids Res. (1988) 16:9347
      13 #title     A major antigen from Mycobacterium tuberculosis which is homologous to the heat shock proteins groES from E. coli and the hsp90 gene product of Coxidia burnetii.
      14 #cross-references MIMD:8401584
      15 #accession S03381
      16 #molecule_type DNA
      17 #residues  1-100 #label BA1
      18 #cross-references EMBL:X12598; NID:944351; PID:9581358
      19 REFERENCE  A37166
      20 #authors   Baird, P.N.; Hall, L.M.C.; Coates, A.R.M.
      21 #journal   J. Gen. Microbiol. (1989) 135:931-939
      22 #title     Cloning and sequence analysis of the 10 kDa antigen gene of Mycobacterium tuberculosis.
      23 #cross-references MIMD:90095443
      24 #accession A37166
      25 #status    preliminary
      26 #molecule_type DNA
      27 #residues  1-100 #label BA2
      28 #cross-references GB:M25258; GB:X12598
      29 REFERENCE  S02727
      30 #authors   Shinnick, T.M.; Plakoytis, B.B.; Hyshe, A.D.; van Landingham, R.M.; Walker, L.L.
      31 #journal   Nucleic Acids Res. (1989) 17:1254
      32 #title     The Mycobacterium tuberculosis BCG-a protein has homology with the Escherichia coli GroES protein.
      33 #cross-references MIMD:89160258

#accession S02727
#molecule_type DNA
#residues 1-100 #label SH1
#cross-references EMBL:X13739; NID:944571; PID:CAA12003.1; PID:9581360
REFERENCE A47292
#authors Keng, I.H.; Coates, A.R.; Butcher, P.D.; Hickman, C.J.; Shinnick, T.M.
#journal Proc. Natl. Acad. Sci. U.S.A. (1993) 90:2608-2612
#title Mycobacterium tuberculosis expresses two chaperonin-40 homologs.
#cross-references MIMD:93219334
#accession A47292
#status preliminary
#molecule_type DNA
#residues 1-100 #label KON
#cross-references EMBL:X60560; NID:944599; PID:CAA12983.1; PID:9581363
#note sequence extracted from NCB: Backbone (NCBI:149955; NCBI:P128607)
REFERENCE A70500
#authors Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, J.; Churcher, C.; Harris, D.; Gordon, S.V.; Eigmeier, K.; Gas, S.; Barry III, C.E.; Tekle, E.; Haddock, K.; Basham, D.; Brown, D.; Chillingworth, T.; Connor, R.; Davies, K.; Devlin, K.; Feltwell, J.; Gentles, S.; Hamlin, N.; Holroyd, N.; Hornsby, J.; Jagels, K.; Krogh, A.; McLean, J.; Miller, S.; Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.; Rajandream, M.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Soares, R.; Sulten, J.; Taylor, K.; Whitehead, S.; Barrett, B.G.
#journal Nature (1998) 393:537-544
#title Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence.
#cross-references MIMD:98295987
#accession G70737
#status nucleic acid sequence not shown; translation not shown
#molecule_type DNA
#residues 1-100 #label GUL
#cross-references GB:Z77165; GB:AI123456; NID:93251609; PID:CA601051; PID:6259180; PID:144969
#exper mental-source strain H37Rv
REFERENCE A46481
#authors Barnes, P.F.; Mehra, V.; Pivoire, B.; Long, S.; Brennan, P.J.; Voegtline, M.S.; Minden, P.; Houghton, R.A.; Bloom, B.R.; Modlin, R.L.
#journal J. Immunol. (1992) 148:1835-1840
#title Immunoreactivity of a 10-kDa antigen of Mycobacterium tuberculosis.
#cross-references MIMD:92176646
#accession A46481
#status preliminary
#molecule_type protein
#residues 231 #label BAW
#note sequence extracted from NCB: Backbone (NCBI:P12128)
GENETICS
#gene groES; cpn10
#start_codon GIG
#complex functional chaperonin includes 14 chains of groEL and 7 of groES
#function mediates protein folding and renaturation
#description #superfamily chaperonin groES
#classification #ATP: heat shock; molecular chaperone; stress-induced protein
#feature 2-100
#product chaperonin groES #status experimental #label MAT
#summary #length 100 #molecular-weight 10804 #checksum 7634
  Query Match 100.0% Score 114; DB 1; Length 100;
  Best Local Similarity 100.0%; Pred No. 1.07e-10;
  Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db      84 EBYLILSARDVLAVSK 100
      1 |||||.....|||

```

QY 1 EYVILSARDVLAVWSK 17

RESULT 4  
ENTRY  
TITLE heat shock protein chpA - Mycobacterium leprae  
ALTERNATE\_NAMES chaperonin\_10K; protein Bl620\_C3\_227  
ORGANISM Mycobacterium leprae  
DATE 19-Mar-1997 #sequence\_revision 25-Apr-1997 #text\_change 19-Mar-1997  
ACCESSIONS S72818  
REFERENCE S72580  
#authors Smith, D.R.; Robison, K.  
#submission Submitted to the EMBL Data Library, November 1993  
#description Mycobacterium leprae cosmid Bl620.  
#accession S72818  
#status preliminary  
#molecule\_type DNA  
#residues 1-169 #label SMI  
#cross-references EXBL:U00015; NID:q466931; PID:AAQ43227.1; PID:q466939

GENETICS  
#gene chpA  
CLASSIFICATION #superfamily chaperonin groES  
KEYWORDS heat shock; molecular chaperone; stress-induced protein  
SUMMARY #length 169 #molecular\_weight 18097 #checksum 6011  
Query Match 100.0%; Score 114; DB 2; Length 169;  
Best Local Similarity 100.0%; Pred. No. 1,07e-10;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

BL 153 EYVILSARDVLAVWSK 169

QY 1 EYVILSARDVLAVWSK 17

RESULT 5  
ENTRY  
TITLE chaperonin groES - Streptomyces coelicolor  
ORGANISM Streptomyces coelicolor  
DATE 05-Jan-1995 #sequence\_revision 06-Jan-1995 #text\_change 26-Aug-1999  
ACCESSIONS S37565  
REFERENCE S37564  
#authors Dubene, A.M.; Kieser, K.; Hopwood, D.; Thompson, C.; Mazodier, P.  
#submission Submitted to the EMBL Data Library, September 1993  
#description Molecular characterization of two groEL genes in Streptomyces coelicolor A3(2).  
#accession S37565  
#status preliminary  
#molecule\_type DNA  
#residues 1-102 #label DCC  
#cross-references EMBL:X75226; NID:q406595; PID:CAA53018.1; PID:q409756

GENETICS  
#gene groES  
#start\_codon GTG  
CLASSIFICATION #superfamily chaperonin groES  
KEYWORDS heat shock; molecular chaperone; stress-induced protein  
SUMMARY #length 102 #molecular\_weight 10945 #checksum 4805

Query Match 92.1%; Score 105; DB 2; Length 102;

Best Local Similarity 82.4%; Pred. No. 1.16e-08;  
Matches 14; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

DB 96 EYVILSARDVLAVWSK 102

QY 1 EYVILSARDVLAVWSK 17

RESULT 6  
ENTRY  
TITLE A41325 #type complete

TITLE heat shock protein 18 - Streptomyces albus  
ALTERNATE\_NAMES heat shock protein groES homolog  
ORGANISM Streptomyces albus  
DATE 17-Jul-1992 #sequence\_revision 17-Jul-1992 #text\_change 12-Sep-1997

ACCESSIONS A41325  
REFERENCE A41325  
#authors Mazodier, P.; Gaglielmi, G.; Davies, J.; Thompson, C.; J. Bacteriol. (1991) 173:7362-7386  
#journal J. Bacteriol. (1991) 173:7362-7386  
#title Characterization of the groEL-like genes in Streptomyces albus.

#cross-references UID:92041639

#accession A41325

#molecule\_type DNA

#residues 1-102 #label MAZ

#cross-references GB:M 57

GENETICS

#start\_codon GTG

CLASSIFICATION #superfamily chaperonin groES

SUMMARY #length 102 #molecular\_weight 10955 #checksum 4443

Query Match 93.4%; Score 103; DB 2; Length 102;

Best Local Similarity 76.5%; Pred. No. 3.24e-08;

Matches 13; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

DB 86 EYVILSARDVLAVWSK 102

QY 1 EYVILSARDVLAVWSK 17

RESULT 7  
ENTRY  
TITLE groES protein - Synechococcus sp. (strain PCC 7942)  
ORGANISM Synechococcus sp.  
DATE 12-Apr-1991 #sequence\_revision 12-Apr-1991 #text\_change 26-Aug-1999  
ACCESSIONS A36721  
REFERENCE A36721  
#authors Webb, R.; Reddy, K.J.; Sherman, L.A.  
#journal J. Bacteriol. (1990) 172:5079-5088  
#title Regulation and sequence of the Synechococcus sp. strain PCC 7942 groES operon, encoding a cyanobacterial chaperonin.

#cross-references UID:9036856;

#accession A36721

#status preliminary

#molecule\_type DNA

#residues 1-103 #label WEB

#cross-references GB:X5870; NID:q154519; PID:AAA27313.1; PID:q154520

CLASSIFICATION #superfamily chaperonin groES

SUMMARY #length 103 #molecular\_weight 10742 #checksum 6554

Query Match 73.2%; Score 80; DB 2; Length 103;

Best Local Similarity 50.0%; Pred. No. 2.39e-04;

Matches 8; Conservative 7; Mismatches 1; Indels 0; Gaps 0;

DB 88 DDVILSEKDLAVWA 103

QY 1 EYVILSARDVLAVWS 16

RESULT 8  
ENTRY  
TITLE BVVCGS #type complete  
ALTERNATE\_NAMES chaperonin groES - Synechococcus sp. (strain PCC 6301)  
ORGANISM Synechococcus sp.  
DATE 30-Sep-1991 #sequence\_revision 30-Sep-1991 #text\_change 16-Jul-1999  
ACCESSIONS S10836; S09619  
REFERENCE S07286  
#authors Cozens, A.L.; Walker, J.E.  
#journal J. Mol. Biol. (1987) 194:359-383  
#title The organization and sequence of the genes for ATP synthase subunits in the cyanobacterium Synechococcus 6301. Support







```

DMPLEX
functional chaperonin includes 14 chains of groEL and 7 of
groES

FUNCTION
  #description: mediates protein folding and renaturation
CLASSIFICATION
  #superfamily: chaperonin groES
KEYWORDS
  ATP; Heat shock; molecular chaperone; stress-induced protein
SUMMARY
  #length 94 #molecular_weight 10176 #checksum 7456

Query Match:
  Query Match: 65.8%; Score 75; DB 2; Length 94;
  Best Local Similarity: 60.0%; Pred. No. 2,30e-02;
  Matches 9; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

DB 80 EYLKRESDIILAVI 94
      1 111 1111111111
QY 2 EYLKRESDIILAVI 15

RESULT 14
ENTRY
  #title: Heat shock protein 10GroES - thermophilic bacterium PS-3
  #ALTERNATE_NAMES: heat shock 10K protein; hsp10
  #ORGANISM: #formal_name: thermophilic bacterium PS-3
  #DATE: 10-Sep-1998; #sequence_revision: 10-Sep-1993; #text_change:
  #ACCESSIONS: J01479; J01494
  #REFERENCE: J01479; J01494
  #authors: Ohta, T.; Honda, K.; Saito, K.; Hayashi, H.; Tano, H.;
    Hamamoto, T.; Kagawa, Y.
  #journal: Biochem. Biophys. Res. Commun. (1993) 191:550-557
  #title: Heat shock promoter of thermophilic chaperonin operon.
  #cross-references: MUID:91213292
  #accession: J01479
  #molecule_type: DNA
  #residues: 1-94; #label: OHT
REFERENCE
  #authors: Yamada, H.; Ohta, T.; Hamamoto, T.; Otawara-Hamamoto, Y.;
    Yada, M.; Hirata, H.; Hirata, H.; Kagawa, Y.
  #journal: Biochem. Biophys. Res. Commun. (1993) 179:565-572
  #title: Gene structure of heat shock proteins 61Kba and 12Kba
    (thermophilic chaperonins) of thermophilic bacterium PS3.
  #cross-references: MUID:91354309
  #accession: J01494
  #molecule_type: DNA
  #residues: 1-94; #label: FAM
  #note: the nucleotide sequence is not complete
GENETICS
  #start_codon: GTG
  #superfamily: chaperonin groES
CLASSIFICATION
  #heat shock: stress-induced protein
KEYWORDS
  #length 94 #molecular_weight 10176 #checksum 7456
SUMMARY

Query Match: 64.9%; Score 74; DB 2; Length 94;
  Best Local Similarity: 64.3%; Pred. No. 3,59e-02;
  Matches 9; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

DB 80 EYLKRESDIILAVI 93
      1 11111 111111
QY 2 EYLKRESDIILAVI 15

RESULT 15
ENTRY
  #title: Heat shock protein GroES - Bacillus sp. (fragment)
  #ALTERNATE_NAMES: molecular chaperone 60 GroES
  #ORGANISM: #formal_name: Bacillus sp.
  #DATE: 20-Feb-1997; #sequence_revision: 27-Feb-1997; #text_change:
    26-Aug-1999
  #ACCESSIONS: PC4238; PC6023
  #REFERENCE: JCS130
  #authors: Xu, Y.; Kobayashi, T.; Kudo, T.
  #journal: Biosci. Biotechnol. Biochem. (1996) 60:1633-1636
  #title: Molecular cloning and nucleotide sequence of the groEL gene
    from the alkaliphilic Bacillus sp. strain C-125 and

```

```

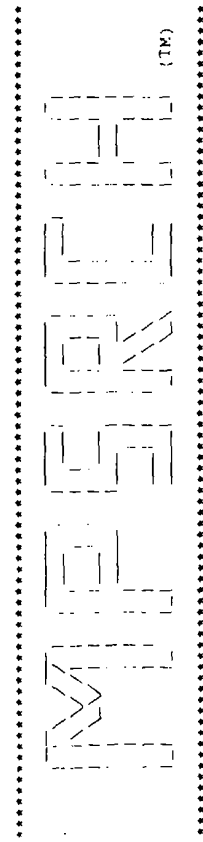
reactivation of thermally inactivated alpha-glucosidase by
recombinant GroEL.
  #cross-references: MUID:97141316
  #accession: PC4238 Preliminary
  #status: Preliminary
  #molecule_type: DNA
  #residues: 1-88; #label: XUA
  #cross-references: DBJ:055610; NID:1664949; PIDN:BAA14941.1;
    PID:0101145; PTC:01682953
  #experimental_source: strain C-125
REFERENCE
  #accession: J06063
  #authors: Xu, Y.; Zhou, P.-J.
  #journal: Acta Microbiol. Sin. (1996) 35:241-249
  #title: Phylogeny of molecular chaperone 60 proteins
  #accession: PC6023
  #molecule_type: DNA
  #residues: 1-88; #label: X02
  #experimental_source: C-125 strain BR224
  #COMMENT: This protein plays a role in protein folding by binding to the
    complex of an unfolding peptide and GroEL protein
GENETICS
  #gene: groES
  #superfamily: chaperonin groES
CLASSIFICATION
  #molecular chaperone
KEYWORDS
  #length 88 #checksum 1072
SUMMARY

Query Match: 64.0%; Score 73; DB 2; Length 88;
  Best Local Similarity: 53.3%; Pred. No. 5,54e-02;
  Matches 8; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

DB 74 EYLKRESDIILAVI 88
      1 11111 11111111
QY 2 EYLKRESDIILAVI 16

Search completed: Mon Jun 19 16:08:09 2000
Job time: 7 secs.

```



Release 3.1A John F. Collins, Bioinformatics Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution Rights by Oxford Molecular Ltd

Search\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Mon Jun 19 16:07:40 2000; Maspar time 3.67 seconds  
Tabular output not generated. 141.095 Million cell updates/sec

Title: XUS-09-142-524A-8  
Description: (1-17) from US09142524A.pep  
Perfect Score: 114  
Sequence: : EYLLSARDVLAVYSK 17

Scoring table: PAM 150  
Gap 15

Search: 8357 seqs, 30454973 residues  
Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: swissprot38  
: swissprot

Statistics: Mean 29.658 Variance 37.117; scale 0.798

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Query | Match | Length | DB | ID         | Description            | Pred. No. |
|------------|-------|-------|--------|----|------------|------------------------|-----------|
| 1          | 114   | 100.0 | 99     | 1  | CH10_MYCTU | 10 KD CHAPERONIN (PROT | 9.22e-13  |
| 2          | 114   | 100.0 | 99     | 1  | CH10_MYCLE | 10 KD CHAPERONIN (PROT | 9.22e-13  |
| 3          | 114   | 100.0 | 99     | 1  | CH10_MYCAV | 10 KD CHAPERONIN (PROT | 9.22e-13  |
| 4          | 105   | 92.1  | 102    | 1  | CH10_STRCO | 10 KD CHAPERONIN (PROT | 1.94e-10  |
| 5          | 103   | 95.4  | 102    | 1  | CH10_STRAL | 10 KD CHAPERONIN (PROT | 5.22e-10  |
| 6          | 80    | 70.2  | 100    | 1  | CH10_THETH | 10 KD CHAPERONIN (PROT | 2.13e-04  |
| 7          | 80    | 70.2  | 103    | 1  | CH10_SYNTP | 10 KD CHAPERONIN (PROT | 2.13e-04  |
| 8          | 80    | 70.2  | 103    | 1  | CH10_SYNTP | 10 KD CHAPERONIN (PROT | 2.13e-04  |
| 9          | 79    | 69.3  | 99     | 1  | CH10_MYCBV | 10 KD CHAPERONIN (PROT | 3.58e-04  |
| 10         | 78    | 68.4  | 102    | 1  | CH10_SYNVU | 10 KD CHAPERONIN (PROT | 6.00e-04  |
| 11         | 78    | 68.4  | 103    | 1  | CH10_CYPAP | 10 KD CHAPERONIN (PROT | 6.00e-04  |
| 12         | 76    | 65.7  | 95     | 1  | CH10_LEPIN | 10 KD CHAPERONIN (PROT | 1.67e-03  |
| 13         | 75    | 65.8  | 94     | 1  | CH10_BAGSU | 10 KD CHAPERONIN (PROT | 2.77e-03  |
| 14         | 75    | 65.8  | 94     | 1  | CH10_BAGP3 | 10 KD CHAPERONIN (PROT | 2.77e-03  |
| 15         | 74    | 64.9  | 94     | 1  | CH10_BACP3 | 10 KD CHAPERONIN (PROT | 4.57e-03  |
| 16         | 74    | 64.9  | 104    | 1  | CH11_BRAJA | 10 KD CHAPERONIN 1 (PR | 4.57e-03  |
| 17         | 73    | 64.0  | 94     | 1  | CH10_THEBR | 10 KD CHAPERONIN (PROT | 7.52e-03  |
| 18         | 71    | 62.3  | 94     | 1  | CH10_LACLA | 10 KD CHAPERONIN (PROT | 2.01e-02  |
| 19         | 70    | 61.4  | 94     | 1  | CH10_STRAC | 10 KD CHAPERONIN (PROT | 3.27e-02  |
| 20         | 70    | 61.4  | 96     | 1  | CH10_HADSD | 10 KD CHAPERONIN (PROT | 3.27e-02  |
| 21         | 59    | 60.5  | 96     | 1  | CH10_AC1PL | 10 KD CHAPERONIN (PROT | 5.23e-02  |
| 22         | 59    | 60.5  | 102    | 1  | CH10_SYNTP | 10 KD CHAPERONIN (PROT | 5.23e-02  |
| 23         | 68    | 59.6  | 96     | 1  | CH10_ACYPS | 10 KD CHAPERONIN (PROT | 8.52e-02  |

|    |    |      |      |   |            |                        |          |
|----|----|------|------|---|------------|------------------------|----------|
| 24 | 67 | 58.8 | 94   | 1 | CH10_CLOTH | 10 KD CHAPERONIN (PROT | 1.37e-01 |
| 25 | 67 | 58.8 | 94   | 1 | CH10_STAEP | 10 KD CHAPERONIN (PROT | 1.37e-01 |
| 26 | 67 | 58.8 | 96   | 1 | CH10_BUCAP | 10 KD CHAPERONIN (PROT | 1.37e-01 |
| 27 | 65 | 57.0 | 94   | 1 | CH10_CLOPE | 10 KD CHAPERONIN (PROT | 3.44e-01 |
| 28 | 65 | 57.0 | 97   | 1 | CH10_VEREN | 10 KD CHAPERONIN (PROT | 3.44e-01 |
| 29 | 65 | 57.0 | 97   | 1 | CH10_SCOLE | 10 KD CHAPERONIN (PROT | 3.44e-01 |
| 30 | 65 | 57.0 | 102  | 1 | CH10_CHLPL | 10 KD CHAPERONIN (PROT | 3.44e-01 |
| 31 | 65 | 57.0 | 255  | 1 | CH10_SPIOL | 20 KD CHAPERONIN (PROT | 3.44e-01 |
| 32 | 64 | 56.1 | 97   | 1 | CH10_XANNA | 10 KD CHAPERONIN (PROT | 5.48e-01 |
| 33 | 64 | 56.1 | 97   | 1 | CH10_PSEST | 10 KD CHAPERONIN (PROT | 5.48e-01 |
| 34 | 63 | 55.3 | 96   | 1 | CH10_HAEIN | 10 KD CHAPERONIN (PROT | 8.51e-01 |
| 35 | 63 | 55.3 | 96   | 1 | CH10_ACIAC | 10 KD CHAPERONIN (PROT | 8.51e-01 |
| 36 | 63 | 55.3 | 96   | 1 | CH10_PASMU | 10 KD CHAPERONIN (PROT | 8.51e-01 |
| 37 | 63 | 55.3 | 96   | 1 | CH10_CXBU  | 10 KD CHAPERONIN (PROT | 8.51e-01 |
| 38 | 63 | 55.3 | 512  | 1 | NUOM_RHOCA | NADH DEHYDROGENASE (C  | 8.51e-01 |
| 39 | 52 | 54.4 | 88   | 1 | CH10_TREPA | 10 KD CHAPERONIN (PROT | 1.35e-00 |
| 40 | 52 | 54.4 | 89   | 1 | CH10_FORGI | 10 KD CHAPERONIN (PROT | 1.35e-00 |
| 41 | 61 | 53.5 | 93   | 1 | CH10_BORBU | 10 KD CHAPERONIN (PROT | 2.10e-00 |
| 42 | 61 | 53.5 | 118  | 1 | CH10_HELPY | 10 KD CHAPERONIN (PROT | 2.10e-00 |
| 43 | 60 | 52.6 | 97   | 1 | CH10_PSEPY | 10 KD CHAPERONIN (PROT | 3.25e-00 |
| 44 | 60 | 52.6 | 332  | 1 | G3P_HELPY  | GLYCERALDEHYDE 3-PHOS  | 3.25e-00 |
| 45 | 60 | 52.6 | 1005 | 1 | EP85_HUMAN | EPH-FAMILY RECEPTOR PR | 3.25e-00 |

ALIGNMENTS

RESULT 1  
ID CH10\_MYCTU STANDARD: PRI: 99 AA.  
AC P09621:  
DI 01-MAR-1989 (Rel. 10. Created;  
D: 15-FEB-2000 (Rel. 39, Last sequence update)  
DE 15-FEB-2000 (Rel. 39, Last annotation update)  
DT 10 KD CHAPERONIN (PROTEIN CPN10) (PROTEIN GROES) (BCG-A HEATSHOCK  
DE PROTEIN) (10 KD ANTIGEN).  
GN MOPB OR GROES OR CPN10 OR RV3418C OR MICV78 11.  
OS Mycobacterium tuberculosis.  
OC Bacteria: Firmicutes; Actinobacteria; Actinobacteridae; Mycobacteriaceae; Mycobacterium.  
CC Actinomycetales; Corynebacteriales; Mycobacteriales; Mycobacterium.  
RN 1:  
RP SEQUENC FROM N.A.  
RC STRAIN=H37RV;  
RX MEDLINE: 89016584.  
RA Baird P.N., Hall L.M., Coates A.R.M.,  
RT "A major antigen from Mycobacterium tuberculosis which is homologous  
RT to the heat shock proteins GroES from E. coli and the hspA gene  
RT product of Coxiella burnetii".  
RL Nucleic Acids Res. 15:9047-9047(1988).  
RN 2:  
RP SEQUENC FROM N.A.  
RC STRAIN=H37RV;  
RX MEDLINE: 89096443.  
RA Baird P.N., Hall L.M., Coates A.R.M.,  
RT "Cloning and sequence analysis of the 10 kDa antigen gene of  
RT Mycobacterium tuberculosis".  
RL J. Gen. Microbiol. 135:931-939(1989).  
RN 3:  
RP SEQUENC FROM N.A.  
RC STRAIN=ERMANN;  
RX MEDLINE: 89160258.  
RA Shinnick T.M., Plikaytis B.P., Hyshe A.D., van Landingham R.M.,  
RT Walker L.L.,  
RT "The Mycobacterium tuberculosis BCG-a protein has homology with the  
RT Escherichia coli GroES protein".  
RL Nucleic Acids Res. 17:1254-1254(1989).  
RN 4:  
RP SEQUENC FROM N.A.  
RC STRAIN=ERMANN;  
RX MEDLINE: 93219332.  
RA Kong I.H., Coates A.R.M., Butcher P.D., Hickman C.J., Shinnick T.M.,  
RT "Mycobacterium tuberculosis expresses two chaperonin-60 homologs".  
RL Proc. Natl. Acad. Sci. U.S.A. 90:2508-2512(1993).  
RN 5:  
RP SEQUENC FROM N.A.



GN MOPR OR GROES.  
 CS Mycobacterium avium, and Mycobacterium paratuberculosis.  
 CC Bacteria: Firmicutes: Actinobacteria: Actinobacteridae;  
 CC Actinomycetales: Corynebacteriineae: Mycobacteriaceae: Mycobacterium.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES-M.AVIUM: STRAIN-485 TYPE 211  
 RA Cresti R., Pietrobono R., Rattorini L., Orefici G.  
 RT "Sequence of the groEL operon of Mycobacterium avium comprising the  
 RT gene encoding the cpn10 protein and a portion of the gene encoding the  
 RI cpn60-1 protein."  
 RE Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES-M.PARATUBERCULOSIS: M.AVIUM: STRAIN-ATCC 19656, ATCC 25291;  
 RA Cobb A.J., Frothingham R.  
 RT "The groES antigens of Mycobacterium avium and Mycobacterium  
 RT paratuberculosis."  
 RE Vet. Microbiol. 0:0-0(1999).  
 CC -1- FUNCTION: BINDS TO CPN60 IN THE PRESENCE OF MG-ATP AND SUPPRESS  
 CC THE ATPASE ACTIVITY OF THE LATTER (BY SIMILARITY).  
 CC -1- SUBUNIT: HEPTAMER OF 7 SUBUNITS ARRANGED IN A RING (BY  
 CC SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE GROES CHAPERONIN FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL: AF079544; AAC31921.1;  
 DR EMBL: AF071829; AAD23277.1;  
 DR EMBL: AF071828; AAD23276.1;  
 DR PFAM: PF00166; cpn10; 1;  
 DR PRINTS: PR00267; CHAPERONIN10;  
 DR PROSITE: PS00681; CHAPERONINS\_CPN10; 1;  
 KW Chaperone; Antigen; Heat shock;  
 KW INIT\_MET C 0 BY SIMILARITY.  
 SV SEQUENCE 99 AA: 10517 MW: 6804437405310C2 CRC64:  
 Query Match 100.0% Score 114; DB 1; Length 99;  
 Best Local Similarity 100.0%; Pred. No. 9,22e-13;  
 Matches 17: Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 DB 83 EBYLVLSARDVLAVSK 49  
 QY 1 EBYLVLSARDVLAVSK 17  
 |||:||||:|||||  
 RESULT 4  
 ID CH10\_STRCO STANDARD: PRI: 102 AA.  
 AC 940772;  
 DT 31-FEB-1995 (Rel. 31; Created);  
 DT 01-FEB-1995 (Rel. 31; Last sequence update);  
 DT 15-DEC-1996 (Rel. 37; Last annotation update)  
 DE 10 KD CHAPERONIN (PROTEIN CPN10) (PROTEIN GROES).  
 GN GROES OR SC624.39  
 OS Streptomyces coelicolor, and Streptomyces lividans.  
 CC Bacteria: Firmicutes: Actinobacteria: Actinobacteridae;  
 CC Actinomycetales: Streptomycetaceae: Streptomyces.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES-S.COELICOLOR: STRAIN-A3(2) / J1501;  
 RX MEDLINE: 94294177.  
 RA Duchene A.M., Kieser H.M., Hopwood D.A., Thompson G.J., Mazodier P.;  
 RT "Characterization of two groEL genes in Streptomyces coelicolor  
 RT A3(2)."  
 RE Gene 144:97-101(1994).  
 RN [2]  
 RP SEQUENCE FROM N.A.

RC SPECIES-S.COELICOLOR: STRAIN-A3(2);  
 RA Saunders D., Harris D., Parkhill J., Barrrell B.G., Ralmandream M.A.;  
 RE Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES-S.LIVIDANS;  
 RX MEDLINE: 98048491;  
 RA de Leon P., Marco S., Islegas C., Marina A., Carrascosa J.L.;  
 RA Mellado R.P.;  
 RT "Streptomyces lividans groES, groEL1 and groEL2 genes."  
 RE Microbiology 143:3563-3571(1997).  
 CC -1- FUNCTION: BINDS TO CPN60 IN THE PRESENCE OF MG-ATP AND SUPPRESSES  
 CC THE ATPASE ACTIVITY OF THE LATTER.  
 CC -1- SUBUNIT: HEPTAMER OF 7 SUBUNITS ARRANGED IN A RING  
 CC (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE GROES CHAPERONIN FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL: X75206; CAA53018.1;  
 DR EMBL: ALC31317; CAA20417.1;  
 DR EMBL: X95970; CAA65224.1;  
 DR PIR: S37565; S37565;  
 DR PRINTS: PR00297; CHAPERONIN10;  
 DR PROSITE: PS00681; CHAPERONINS\_CPN10; 1;  
 DR PFAM: PF00166; cpn10; 1;  
 KW Chaperone.  
 SV SEQUENCE 102 AA: 10946 MW: 60F9FA68014AC8CC CRC64:  
 Query Match 92.1% Score 105; DB 1; Length 102;  
 Best Local Similarity 82.4%; Pred. No. 1,94e-19;  
 Matches 14: Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 DB 86 EBYLVLSARDVLAVSK 102  
 QY 1 EBYLVLSARDVLAVSK 17  
 |||:||||:|||||  
 RESULT 5  
 ID CH10\_STRAL STANDARD: PRI: 102 AA.  
 AC Q00769;  
 DT 01-APR-1993 (Rel. 25; Created);  
 DT 01-APR-1993 (Rel. 25; Last sequence update);  
 DT 01-FEB-1996 (Rel. 33; Last annotation update);  
 DE 10 KD CHAPERONIN (PROTEIN CPN10) (PROTEIN GROES).  
 GN GROES  
 OS Streptomyces albus J.  
 CC Bacteria: Firmicutes: Actinobacteria: Actinobacteridae;  
 CC Actinomycetales: Streptomycetaceae: Streptomycetaceae: Streptomyces.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 92041639.  
 RA Mazodier P., Gadhielemi G., Davies J., Thompson G.J.;  
 RT "Characterization of the groEL-like genes in Streptomyces albus J."  
 RE Bacteriol. 173:7382-7386(1994).  
 CC -1- FUNCTION: BINDS TO CPN60 IN THE PRESENCE OF MG-ATP AND SUPPRESSES  
 CC THE ATPASE ACTIVITY OF THE LATTER.  
 CC -1- SUBUNIT: HEPTAMER OF 7 SUBUNITS ARRANGED IN A RING  
 CC (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE GROES CHAPERONIN FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>



```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: X13970; CAA32149.1; -
CC DR EMBL: M35389; AAA26365.1; -
CC DR PIR: S0784; BVMY7B.
CC DR PRINTS: PR00297; CHAPERONIN.O.
CC DR PROSITE: PS00681; CHAPERONINS_CPN10.1.
CC DR PFAM: PF00166; cpn10.1.
CC DR Chaperone; Antigen; Heat shock.
CC KW
CC FT INIL MET
CC FT INIL MET
CC SEQUENCE 99 AA: 108-9 MW: 100552199447AF7 CRC64:
CC
CC
CC Query Match 69.3% Score 79; DB 1; Length 99;
CC Best Local Similarity 84.6%; Pred. No. 3,56e-04;
CC Matches 11; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
CC
CC Db 63 EBYLLSARDVVG 95
CC QY 1 EBYLLSARDVJA 13
CC
CC RESULT 10 STANDARD; PRI: 102 AA.
CC ID CH10.SYNVU
CC AC Q50322.
CC DT 15-DEC-1998 (Rel. 37, Created;
CC DT 15-DEC-1998 (Rel. 37, Last sequence update)
CC DT 15-DEC-1998 (Rel. 37, Last annotation update)
CC DE 10 K3 CHAPERONIN (PROTEIN CPN10) (PROTEIN GROES).
CC GN GROES.
CC OS Synecococcus vulcanus.
CC OC Bacteria; Cyanobacteria; Chroococcales; Synecococcus.
CC RN
CC RP SEQUENCE FROM N.A.
CC RX MEDLINE; 88094212.
CC RA Tanaka N., Hiyma T., Nakamoto H.
CC RT "Cloning, characterization and functional analysis of grcFSL gene
CC RT from thermophilic cyanobacterium Synechococcus vulcanus."
CC FL Biochim. Biophys. Acta 1343:335-348(1997).
CC CC -F- FUNCTION: BINDS TO CPN50 IN THE PRESENCE OF MG-ATP AND SUPPRESSES
CC CC THE ATPASE ACTIVITY OF THE LATTER.
CC CC -S- SUPUNIT: HEPTAMER OF 7 SUBUNITS ARRANGED IN A RING
CC CC -S- SYMILARITY BELONGS TO THE GROES "HAPD" AND FAMILY
CC CC This SWISS-PROT entry is copyrighted. It is provided through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: D78139; BAA23816.1; -
CC DR PRINTS: PR00297; CHAPERONIN.O.
CC DR PROSITE: PS00681; CHAPERONINS_CPN10.1.
CC DR PFAM: PF00166; cpn10.1.
CC DR Chaperone.
CC KW
CC FT INIL MET
CC FT INIL MET
CC SEQUENCE 102 AA: 10783 MW: 87003617298BFF67 CRC64:
CC
CC
CC Query Match 68.4% Score 78; DB 1; Length 102;
CC Best Local Similarity 50.0%; Pred. No. 6,00e-04;
CC Matches 8; Conservative 7; Mismatches 1; Indels 0; Gaps 0;
CC
CC Db 87 EDYVLLSEKDLIAVG 102
CC QY 1 EBYLLSARDVLAIVS 16

```

```

RESULT 11
ID CHIC CYAPA STANDARD: PSI: 103 AA.
AC Q3762
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE 10 KD CHAPERONIN (PROTEIN CN10) (PROTEIN GROES)
GN GROES-A AND GROES-B
OS Cyanochoira paradoxa
OC Cyanochoira
OC Eukaryota: Glomerophyceae, Cyanochoiraceae: Cyanochoira
RN [1]
RF SEQUENCE FROM N.A.
RC STRAIN L555 / PR-NGSHEW
RA Sullivan V.L., Michalowski C.B., Lottelhard W., Bohnert H.J.,
RA Bryant J.A.
RL Submitted (OCT-1975) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: BINDS TO CPN60 IN THE PRESENCE OF MG-ATP AND SUPPRESSES
CC THE ATPASE ACTIVITY OF THE LATTER (BY SIMILARITY).
CC -1- SUBUNIT: HEPTAMER OF 7 SUBUNITS ARRANGED IN A RING
CC (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE GROES CHAPERONIN FAMILY.
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U30821; AAA81731.1
CC EMBL: U30821; AAA82921.1
CC PRIN: PRO0297; CHAPERONIN10
CC PROSITE: PS00681; CHAPERONINS_CN10_1
CC PFAM: PF00166; cpn10_1
CC Chaperone: Cyanochoira
KW Chaperone
SQ SEQUENCE 103 AA: 11240 MW: 12240 MW: FF694790CA90520 GR*64:
Query Match 68.48: Score 78: DB 1: Length 103:
Best Local Similarity 50.00: Pred. No. 6,00e-04:
Matches 8: Conservative 7: Mismatches 1: Indels 0: Gaps 0:
DB 88 REVVLSEXOILALIA 103
ID REVVLSEXOILALIA 103
AC REVLSAROVLAWS 16
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE 10 KD CHAPERONIN (PROTEIN CN10) (PROTEIN GROES) (HEAT SHOCK 10 KD
DE PROTEIN)
GN GROES OR HSP10
OS Leptospira interrogans
OC Bacteria: Spirochaetales: Leptospiraceae: Leptospira
RN [1]
RF SEQUENCE FROM N.A.
RC STRAIN-SEROVAR COPENHAGEN1 / WIJNBORG
RA MEDLINE: 93423752.
RA Ballard S.A., Sedgers R.P., Blomnick-Pluym N.M., Fyfe J.A.M.,
RA Faine S., Adler B.
RT "Molecular analysis of the hsp (groE) operon of Leptospira
RT interrogans serovar Copenhagen1."
RL Mol. Microbiol. 8:739-751(1993).
RN [2]
RF SEQUENCE FROM N.A.
RC STRAIN-SEROVAR DAI / HY-1
RA Kim M.J., Ahn B.Y.

```

```

RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: BINDS TO CPN60 IN THE PRESENCE OF MG-ATP AND SUPPRESSES
CC THE ATPASE ACTIVITY OF THE LATTER.
CC -1- SUBUNIT: HEPTAMER OF 7 SUBUNITS ARRANGED IN A RING
CC (BY SIMILARITY).
CC -1- INDUCTION: BY HEAT SHOCK.
CC -1- SIMILARITY: BELONGS TO THE GROES CHAPERONIN FAMILY.
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U34502; AAA71991.1
CC EMBL: AF032910; AAB95964.1
CC PR: S34537; S34937.
CC PRIN: PRO0297; CHAPERONIN10
CC PROSITE: PS00681; CHAPERONINS_CN10_1
CC PFAM: PF00166; cpn10_1
CC Chaperone: Heat shock.
KW Chaperone
SQ SEQUENCE 96 AA: 10562 MW: 925087989F1467 GR*54:
Query Match 66.78: Score 76: DB 1: Length 96:
Best Local Similarity 52.58: Pred. No. 1.57e-03:
Matches 10: Conservative 2: Mismatches 4: Indels 0: Gaps 0:
DB 81 EYLRESIDILAVVK 96
ID EYLRESIDILAVVK 96
AC EYLISAROVLAWSK 17
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE 10 KD CHAPERONIN (PROTEIN CN10) (PROTEIN GROES)
GN GROES
OS Bacillus subtilis.
OC Bacteria: Firmicutes: Bacillus/Clostridium group:
OC Bacillus/Staphylococcus group: Bacillus.
RN [1]
RF SEQUENCE FROM N.A.
RC MEDLINE: 92283754.
RA Schmidt A., Scherzinger M., Winkler F., Boeser M., Schmitt N.,
RA "Cloning, sequencing, mapping, and transcriptional analysis of the
RA groES operon from Bacillus subtilis."
RL J. Bacteriol. 174:1991-1994(1992).
RN [2]
RF SEQUENCE FROM N.A.
RC STRAIN-168 / MARRBURG
RA MEDLINE: 93129852.
RA Tozawa Y., Yoshikawa H., Kawamura F., Itaya M., Takahashi H.
RT "Isolation and characterization of the groES and groEL genes of
RT Bacillus subtilis Marburg."
RL Biosci. Biotechnol. Biochem. 56:1995-2002(1992).
RN [4]
RF SEQUENCE FROM N.A.
RC STRAIN-168 / MARRBURG
RA MEDLINE: 97345038.
RA Sadale Y., Yata K., Fujita M., Sazai H., Itaya M., Kasahara Y.

```

```

RX  MECLINE: 93224474.
RA  Schoen U., Schumann W.:
   "Molecular cloning, sequencing, and transcriptional analysis of the
   groES operon from Bacillus stearothermophilus."
RL  J. Bacteriol. 175:2455-2459(1993).
CC  -!- FUNCTION: BINDS TO CPN60 IN THE PRESENCE OF MG-ATP AND SUPPRESSES
CC  THE ATPASE ACTIVITY OF THE LATTER.
CC  -!- SUBUNIT: HEPTAMER OF 7 SUBUNITS ARRANGED IN A RING
CC  (BY SIMILARITY).
CC  -!- INDUCTION: BY HEAT SHOCK.
CC  -!- SIMILARITY: BELONGS TO THE GROES CHAPERONIN FAMILY.
CC  -----
CC  THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is free of any
CC  modification and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/)
CC  or send an email to license@isb-sib.ch.
CC  -----
CC  EMBL: L01332; AAA22751.1; --
DR  PRUNTS; P800297; CHAPERONIN10.
DR  PROSITE: P500681; CHAPERONINS_CPN10; 1.
DR  Pfam: PF00166; cpn10; 1.
DR  Chaperone: Heat shock.
KW  Chaperone.
SQ  SEQUENCE 94 AA: 1024; MW: 85732A0C42114921 CRC64:
      Query Match      65.8%  Score 75:  DB 1: Length 94.
      Best Local Similarity 60.0%  Pred. No. 2 77e-03
      Matches 9:  Conservative 1:  Mismatches 3:  Indels 0:  Gaps 0:
Db  80 EYLILRESDILAVIG 94
      |||||  |||||
QY  2 EYLILSARDVLAVS 16
      |||||  |||||

RESULT 15
ID  CH10_BACP3  STANDARD:  PRT:  94 AA.
AC  P26210;
DT  01-MAY-1992 (Rel. 22. C *ed)
DT  01-MAY-1992 (Rel. 22. Ls sequence update)
DT  01-NOV-1997 (Rel. 35. Last annotation update)
DE  10 KD CHAPERONIN (PROTEIN CPN10) (PROTEIN GROES) (HEAT SHOCK 12 KD.
DE  PROTEIN).
DE  GROES.
CS  Bacillus PS3 (Thermophilic bacterium PS-3).
CC  Bacteria: Firmicutes; Bacillus/Clostridium group.
CC  Bacillus/staphylococcus group; Bacillus.
RN  11.
RP  SEQUENCE FROM N.A.
RX  MEDLINE: 91354309.
RA  Tamada H., Ohta T., Hamamoto T., Mikawa H., Hamamoto Y., Yashii M.,
RA  Hiraiwa H., Hirata H., Kazawa Y.
RT  "Gene structure of heat shock proteins 61kDa and 12kDa (thermophilic
RL  chaperonins) of thermophilic bacterium PS3."
RI  Biochem. Biophys. Res. Commun. 179:565-571(1991).
RN  [2].
RP  SEQUENCE FROM N.A.
RX  MEDLINE: 93213292.
RA  Ohta T., Honda K., Saito K., Hayashi H., Tano H., Hamamoto Y.
RA  Kazawa Y.
RT  "Heat shock promoter of thermophilic chaperonin operon."
RI  Biochem. Biophys. Res. Commun. 191:550-557(1993).
CC  -!- FUNCTION: BINDS TO CPN50 IN THE PRESENCE OF MG-ATP AND SUPPRESSES
CC  THE ATPASE ACTIVITY OF THE LATTER.
CC  -!- SUBUNIT: HEPTAMER OF 7 SUBUNITS ARRANGED IN A RING
CC  (BY SIMILARITY).
CC  -!- SIMILARITY: BELONGS TO THE GROES CHAPERONIN FAMILY.
CC  -----
CC  THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is free of any

```



CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch))  
 CC .....

DR EMBL: S57424; CAB31564.1; ALT\_SEQ.  
 DR PIR: J01194; J01194.  
 DR PIR: J01479; J01479.  
 DR PRINTS: PR00297; CHAPERONIN10.  
 DR PROSITE: PS00681; CHAPERONINS\_CPN10; 1.  
 DR PFAM: PF00166; cpn10; 1.  
 DR Chaperone.  
 KK Chaperone.  
 SC SEQUENCE 94 AA: 10340 MW: 24944954AE0B457C CRC64:

Query Match 54.9% Score 74; DB 1; Length 94;  
 Best Local Similarity 64.3% Pred. No. 4.57e-03;  
 Matches 9; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

DR 80 EYLILRESILAV: 94  
 QY 2 EYLILSARGVLAV: 15

Search completed: Mon Jun 19 16:07:46 2000  
 CPU time: 6 secs.

\*\*\*\*\*  
WATERMANN  
\*\*\*\*\*  
(TM)

Release 3.1A John F. Collins, BioComputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MFsrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Mon Jun 19 16:07:11 2000 Maspar time 8.85 Seconds  
133.146 Million cell updates/sec

Tabular output not generated.

Title: >US-09-142-524A-8  
Description: (1-17) from US09142524A.pep  
Perfect Score: 114  
Sequence: 1 EEVL12SARCVLAVVSK 17

Scoring table: PAM 150  
Gap 15

Searched: 225878 seqs. 69334122 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: spiremb12

1:sp:archaea 2:sp:bacteria 3:sp:functi 4:sp:human  
5:sp:invertebrate 6:sp:mammal 7:sp:mhc 8:sp:organelle  
9:sp:phase 10:sp:plant 11:sp:rodent 12:sp:unclassified  
13:sp:vertebrate 14:sp:vi us

Statistics: Mean 28.205; Variance 39.50; scale 0.713

Pred. No. is the number of results predicted by chance to have a  
Score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description                   | Pred. No. |
|------------|-------|-------------|--------|-------|-------------------------------|-----------|
| 1          | 114   | 100.0       | 100    | 2     | 086017 10 KD CHAPERONIN (PROT | 2.76e-11  |
| 2          | 76    | 65.7        | 103    | 2     | 09X206 10 KD CHAPERONIN (PROT | 8.15e-03  |
| 3          | 73    | 64.0        | 88     | 2     | 050324 GROES (FRAGMENT)       | 3.27e-02  |
| 4          | 70    | 61.4        | 100    | 2     | 09XCBC 10 KD CHAPERONIN (PROT | 1.27e-01  |
| 5          | 68    | 59.6        | 94     | 2     | 068323 10 KD CHAPERONIN (PROT | 3.10e-01  |
| 6          | 67    | 58.8        | 93     | 2     | 032846 10 KD CHAPERONIN (PROT | 4.81e-01  |
| 7          | 67    | 58.8        | 122    | 2     | 067942 GROES                  | 4.81e-01  |
| 8          | 67    | 58.8        | 304    | 13    | Q9W534 MRNA, PARTIAL CDS, SIM | 4.81e-01  |
| 9          | 67    | 59.8        | 312    | 13    | Q9W525 XIMPACT                | 4.81e-01  |
| 10         | 66    | 57.9        | 455    | 4     | 094899 KIAA0795 PROTEIN (FRAG | 7.43e-01  |
| 11         | 65    | 57.0        | 88     | 2     | 066201 SIMILAR TO GROES PROTE | 1.14e-00  |
| 12         | 65    | 57.0        | 88     | 2     | 066189 SIMILAR TO GROES PROTE | 1.14e-00  |
| 13         | 65    | 57.0        | 88     | 2     | 066211 SIMILAR TO GROES PROTE | 1.14e-00  |
| 14         | 65    | 57.0        | 88     | 2     | 066193 SIMILAR TO GROES PROTE | 1.14e-00  |
| 15         | 65    | 57.0        | 88     | 2     | 066219 SIMILAR TO GROES PROTE | 1.14e-00  |
| 16         | 65    | 57.0        | 88     | 2     | 066203 SIMILAR TO GROES PROTE | 1.14e-00  |
| 17         | 65    | 57.0        | 88     | 2     | 066215 SIMILAR TO GROES PROTE | 1.14e-00  |
| 18         | 65    | 57.0        | 88     | 2     | 066205 SIMILAR TO GROES PROTE | 1.14e-00  |
| 19         | 65    | 57.0        | 88     | 2     | 066195 SIMILAR TO GROES PROTE | 1.14e-00  |
| 20         | 65    | 57.0        | 88     | 2     | 066193 SIMILAR TO GROES PROTE | 1.14e-00  |

|    |    |      |     |    |        |                        |          |
|----|----|------|-----|----|--------|------------------------|----------|
| 21 | 65 | 57.0 | 88  | 2  | 086207 | SIMILAR TO GROES PROTE | 1.14e-00 |
| 22 | 65 | 57.0 | 88  | 2  | 05213  | SIMILAR TO GROES PROTE | 1.14e-00 |
| 23 | 65 | 57.0 | 88  | 2  | 0209   | SIMILAR TO GROES PROTE | 1.14e-00 |
| 24 | 65 | 57.0 | 88  | 2  | 0197   | SIMILAR TO GROES PROTE | 1.14e-00 |
| 25 | 65 | 57.0 | 88  | 2  | 0191   | SIMILAR TO GROES PROTE | 1.14e-00 |
| 26 | 65 | 57.0 | 154 | 2  | 111    | HYPOTHEICAL 17.8 KDa P | 1.14e-00 |
| 27 | 64 | 56.1 | 89  | 2  | 05487  | GROES (FRAGMENT)       | 1.14e-00 |
| 28 | 64 | 56.1 | 97  | 2  | 068308 | 10 KD CHAPERONIN (PROT | 1.14e-00 |
| 29 | 63 | 55.3 | 88  | 2  | 045217 | SIMILAR TO GROES PROTE | 2.4e-00  |
| 30 | 63 | 55.3 | 88  | 2  | 066221 | SIMILAR TO GROES PROTE | 2.5e-00  |
| 31 | 63 | 55.3 | 96  | 2  | 051831 | 10 KD CHAPERONIN (PROT | 2.5e-00  |
| 32 | 63 | 55.3 | 261 | 3  | 042628 | 32 KDa-CELL WALL SYMB  | 2.45e-00 |
| 33 | 63 | 55.3 | 330 | 8  | 092203 | NADH-UBIQUINONE OXIDOR | 2.67e-00 |
| 34 | 63 | 55.3 | 445 | 2  | 04390  | NEGATIVE ALPHATIC AM   | 2.67e-00 |
| 35 | 63 | 55.3 | 629 | 10 | 01274  | POTATIVE ALCOHOL DEHYD | 2.67e-00 |
| 36 | 63 | 55.3 | 302 | 4  | 01485  | X-LIKE 1 PROTEIN       | 2.67e-00 |
| 37 | 62 | 54.4 | 305 | 1  | 01495  | HYPOTHEICAL 14.8 KDa P | 2.67e-00 |
| 38 | 61 | 53.5 | 114 | 5  | 063    | HEMOLYSIN              | 6.14e-00 |
| 39 | 61 | 53.5 | 206 | 2  | 065    | THYMIDYLATE KINASE     | 6.14e-00 |
| 40 | 61 | 53.5 | 242 | 5  | 136    | ERCC1-LIKE EXCISION RE | 6.14e-00 |
| 41 | 61 | 53.5 | 300 | 5  | 01222  | SERINE/THREONINE PROT  | 6.14e-00 |
| 42 | 61 | 53.5 | 305 | 5  | 01239  | CHK-RELATED PROTEIN K1 | 6.14e-00 |
| 43 | 61 | 53.5 | 370 | 5  | 026311 | COSCE12 PROTEIN        | 6.14e-00 |
| 44 | 61 | 53.5 | 387 | 1  | 095884 | CRP C02004             | 6.14e-00 |
| 45 | 61 | 53.5 | 476 | 2  | 09X805 | POTATIVE AMINOTRANSFER | 6.14e-00 |

ALIGNMENTS

RESULT 1  
ID 086017 PRELIMINARY: PRT: 100 AA.  
AC 086017:  
D7 C1-NOV-1998 (TREMblrel. 08. Created)  
D7 C1-NOV-1998 (TREMblrel. 08. Last sequence update)  
D7 C1-NOV-1998 (TREMblrel. 12. Last annotation update)  
DE 10 KD CHAPERONIN (PROTEIN CPN10) (PROTEIN GROES).  
GN GROES.  
OS Mycobacterium avium, and Mycobacterium paratuberculosis.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
RN [1]  
PP SEQUENCE FROM N.A.  
RC STRAIN-495 TYPE 21:  
RA CREIG R., PETROBOBO R., PATRINI L., GREFFICI G.  
RT "Sequence of the GroES operon of Mycobacterium avium comprising the  
RT gene encoding the cpn10 protein and a portion of the gene encoding the  
RT cpn60-1 protein."  
RL Submitted (JUL-1999) to the EMBL/GenBank/CCDS databases.  
RN [2]  
PP SEQUENCE FROM N.A.  
RC SPECIES: PARATUBERCULOSIS, STRAIN-AV111 (1994, A111-256)  
RA COBB A.C., PROTHINGHAM R.  
RT "The GroES antigens of Mycobacterium avium and Mycobacterium  
RL paratuberculosis."  
VE: Microbiol. 0:0-0(1999).  
CC -- FUNCTION: BINDS TO CPN60 IN THE PRESENCE OF MG-AIP AND SUPPRESSES  
CC THE ATPASE ACTIVITY OF THE LATTER (BY SIMILARITY).  
CC -- SUBUNIT: HEPTAMER OF 7 SUBUNITS ARRANGED IN A RING (BY  
CC SIMILARITY).  
CC -- SIMILARITY: BELONGS TO THE GROES CHAPERONIN FAMILY.  
DR EMBL: AF079544; AAC31921.1; .  
DR EMBL: AF071829; AAD23277.1; .  
DR EMBL: AF071828; AAD23276.1; .  
DR PROSITE: PS00681; CHAPERONINS\_CPN10; 1.  
DR PFM: PF00166; cpn10; 1.  
DR PRINTS: PR00297; CHAPERON N10.  
KW Chaperone; Heat shock.  
SQ SEQUENCE 100 AA: 10748 M 0C4954DA CRC32:  
Query Match 100.0% score 114; DB 2; Length 100;  
Best Local Similarity 100.0%; Pred.No. 2.76e-11;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;



Q: 1 EYLILSARDVLAV 14

RESULT 6  
ID Q32846 PRELIMINARY: PRT: 93 AA.  
AC Q32846  
DI 01-JAN-1998 (TRENBLrel. 05, Created)  
DI 01-JAN-1998 (TRENBLrel. 05, Last sequence update)  
DI 01-NOV-1999 (TRENBLrel. 12, Last annotation update)  
DI 10 KD CHAPERONIN (PROTEIN CPN10) (PROTEIN GROES)  
GN GROES  
OS Lactobacillus zeae.  
OT Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;  
OL Lactobacillus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RT STRAIN-1025 (PLASMID-FREE DERIVATIVE - ATCC39393)  
RA MURPHY C.M., CHASSY B.M.;  
RC Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.  
CC -- FUNCTION: BINDS TO CPN60 IN THE PRESENCE OF MG-ATP AND SUPPRESS  
OT THE ATPASE ACTIVITY OF THE LATTER (BY SIMILARITY).  
CC -- SUBUNIT: REPETIT OF 7 SUBUNITS ARRANGED IN A RING (BY  
OT SIMILARITY).  
CC -- SIMILARITY: BELONGS TO THE GROES CHAPERONIN FAMILY.  
CP EMBL: AF010251; AAB56325.1;  
DR PROSITE: P500691; CHAPERONINS\_CPNI0; 1.  
DR PFAM: PF00165; CPN10; 1.  
DR PRINTS: PR02977; CHAPERONIN10.  
KA Chaperone.  
SQ SEQUENCE 93 AA: 10040 MW: 35954571 CRC32:

Query Match 58.8% Score 67; DB 2; Length 93;  
Best Local Similarity 35.7% Pred. No. 4,81e-01;  
Matches 5; Conservative 7; Mismatches 2; Indels 0; Gaps 0;

DB 79 CDYLVEHEKDMAI 92  
Q: 1 EYLILSARDVLAV 14

RESULT 7  
ID Q67942 PRELIMINARY: PRT: 122 AA.  
AC Q67942  
DI 01-AGS-1998 (TRENBLrel. 07, Created)  
DI 01-AGS-1998 (TRENBLrel. 07, Last sequence update)  
DI 01-NOV-1999 (TRENBLrel. 12, Last annotation update)  
GN GROES  
GN MOPR.  
OS Aquifex aeolicus.  
OT Bacteria; Aquificales; Aquificaceae; Aquifex.  
RN [1]  
RP SEQUENCE FROM N.A.  
RT STRAIN-VF5.  
RX MEDLINE: 98155666.  
RA DECKERT G., WARREN P.V., GAASTERLAND T., YOUNG W.G., LENOX A.L.,  
RA GRAHAM D.E., OVERBEEK R., SNEAD M.A., KELLER M., AUJAY M., HUBER R.,  
RA FELDMAN R.A., SHORT J.M., OLSON G.J., SWANSON R.V.;  
RC Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.  
RI The complete genome of the hyperthermophilic bacterium Aquifex  
OT aeolicus.  
RL Nature 392:353-358(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RT STRAIN-VF5.  
RX MEDLINE: 98155666.  
RA DECKERT G., WARREN P.V., GAASTERLAND T., YOUNG W.G., LENOX A.L.,  
RA GRAHAM D.E., OVERBEEK R., SNEAD M.A., KELLER M., AUJAY M., HUBER R.,  
RA FELDMAN R.A., SHORT J.M., OLSON G.J., SWANSON R.V.;  
RC Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.  
DI 10 KD CHAPERONIN (PROTEIN CPN10) (PROTEIN GROES)  
GN GROES  
OS Lactobacillus zeae.  
OT Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;  
OL Lactobacillus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RT STRAIN-1025 (PLASMID-FREE DERIVATIVE - ATCC39393)  
RA MURPHY C.M., CHASSY B.M.;  
RC Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.  
CC -- FUNCTION: BINDS TO CPN60 IN THE PRESENCE OF MG-ATP AND SUPPRESS  
OT THE ATPASE ACTIVITY OF THE LATTER (BY SIMILARITY).  
CC -- SUBUNIT: REPETIT OF 7 SUBUNITS ARRANGED IN A RING (BY  
OT SIMILARITY).  
CC -- SIMILARITY: BELONGS TO THE GROES CHAPERONIN FAMILY.  
CP EMBL: AF010251; AAB56325.1;  
DR PROSITE: P500691; CHAPERONINS\_CPNI0; 1.  
DR PFAM: PF00165; CPN10; 1.  
DR PRINTS: PR02977; CHAPERONIN10.  
KA Chaperone.  
SQ SEQUENCE 93 AA: 10040 MW: 35954571 CRC32:

Query Match 58.8% Score 67; DB 2; Length 122;  
Best Local Similarity 35.7% Pred. No. 4,81e-01;  
Matches 5; Conservative 7; Mismatches 2; Indels 0; Gaps 0;

Best Local Similarity 61.5% Pred. No. 4,81e-01;  
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

DB 83 YLWSEDEVCAVV 95  
Q: 3 YLILSARDVLAV 15  
RESULT 8  
ID Q9W634 PRELIMINARY: PRT: 304 AA.  
AC Q9W634  
DI 01-NOV-1999 (TRENBLrel. 12, Created)  
DI 01-NOV-1999 (TRENBLrel. 12, Last sequence update)  
DI 01-NOV-1999 (TRENBLrel. 12, Last annotation update)  
DE MRNA, PARTIAL CDS, SIMILAR TO MOUSE IMPRINTED  
DE GENE IMPACT (FRAGMENT).  
OS Xenopus borealis (Kenyan clawed frog).  
OT Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Amphibia;  
OC Batrachia; Anura; Mesobatrachia; Pipiloidea; Pipidae; Xenopodinae;  
OC Xenopus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA YAMADA Y.;  
RT "Xenopus borealis homologue of mouse imprinted gene impact."  
RC Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AB021258; BAA78421.1;  
FT NON-TER;  
FT NON-TER 304 304  
SQ SEQUENCE 304 AA: 34661 MW: D94F43B1 CRC32:

Query Match 58.8% Score 67; DB 13; Length 304;  
Best Local Similarity 69.2% Pred. No. 4,81e-01;  
Matches 9; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

DB 248 ILDARDVMVYVS 260  
Q: 5 ILSARDVLAVSK 17  
RESULT 9  
ID Q9W625 PRELIMINARY: PRT: 312 AA.  
AC Q9W625  
DI 01-NOV-1999 (TRENBLrel. 12, Created)  
DI 01-NOV-1999 (TRENBLrel. 12, Last sequence update)  
DI 01-NOV-1999 (TRENBLrel. 12, Last annotation update)  
DE XIMPACT.  
OS Xenopus laevis (African clawed frog).  
OT Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Amphibia;  
OC Batrachia; Anura; Mesobatrachia; Pipiloidea; Pipidae; Xenopodinae;  
OC Xenopus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 99167356.  
RA YAMADA Y., HAGIWARA Y., SHICKAWA K., SAKAKI Y., ITO T.;  
RT "Spatialtemporal, allelic, and enforced expression of Ximpact, the  
RI Xenopus homolog of mouse imprinted gene impact."  
RL Biochem. Biophys. Res. Commun. 258:162-169(1999).  
DR EMBL: AB020319; BAA78409.1;  
SQ SEQUENCE 312 AA: 35640 MW: QC25D754 CRC32:

Query Match 58.8% Score 67; DB 13; Length 312;  
Best Local Similarity 69.2% Pred. No. 4,81e-01;  
Matches 9; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

DB 251 ILDARDVMVYVS 263  
Q: 5 ILSARDVLAVSK 17

RESULT 10  
ID Q94889 PRELIMINARY: PRT: 465 AA.  
AC Q94889  
DI 01-MAY-1999 (TRENBLrel. 10, Created)

```

DT 01-MAY-1999 (TREMBLER: 10, Last sequence update)
DI 01-MAY-1999 (TREMBLER: 10, Last annotation update)
DE KIAA0745 PROTEIN (FRAGMENT)
GN KIAA0745
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Euteria; Primates; Catarrhini; Homidae; Homo
RN [1]
RT SEQUENCE FROM N.A.
RC TISSUE-BRAIN
RA MEDLINE: 9907487.
RA NAGASE T., ISHIKAWA K., SUYAMA M., KIKUNO R., MIYAJIMA N., TANAKA A.,
RA KOTANI H., NOMURA N., GHARA G.
RA *Prediction of the coding sequences of unidentified human genes. XI.
RA The complete sequences of 100 new cDNA clones from brain which code
RA for large proteins in vitro.
RL CNA Res. 5:277-286(1998).
DR EMBL: AB018399; BAA34151.1; -.
FT NON-TER
SQ SEQUENCE 465 AA: 51392 MW: 55192666 CRC32:

Query Match 57.9% Score 66: DB 4: Length 465:
Best Local Similarity 60.0% Pred. No. 1,140-00:
Matches 9: Conservative 4: Mismatches 3: Indels 0: Gaps 0:

DB 65 EBYLMSRDLVLSR 86
QY 1 EBYLMSRDLVLSR 15

RESULT 1:
ID 066201 PRELIMINARY: PRT: 88 AA
AC 066201
DI 01-AUG-1998 (TREMBLER: 07, Created)
DI 01-AUG-1998 (TREMBLER: 07, Last sequence update)
DE 01-MAY-1999 (TREMBLER: 10, Last annotation update)
DE SIMILAR TO GROS PROTEIN (FRAGMENT)
OS Seriatia rubidacea
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Seriatia
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-JCM 1240:
RA HARADA H., ISHIKAWA H.
RA J. Gen. Appl. Microbiol. 43:355-361(1997).
DR EMBL: AB008143; BAA25218.1; -.
DR PFAM: PF00166; cpn10.1.
FT NON-TER
SQ SEQUENCE 84 AA: 9212 MW: 25199566 CRC32:

Query Match 57.9% Score 65: DB 2: Length 86:
Best Local Similarity 60.0% Pred. No. 1,140-00:
Matches 9: Conservative 3: Mismatches 3: Indels 0: Gaps 0:

DB 72 EBYLMSRDLVLSR 86
QY 1 EBYLMSRDLVLSR 15

RESULT 12
ID 066189 PRELIMINARY: PRT: 88 AA
AC 066189
DI 01-AUG-1998 (TREMBLER: 07, Created)
DI 01-AUG-1998 (TREMBLER: 07, Last sequence update)
DE 01-MAY-1999 (TREMBLER: 10, Last annotation update)
DE SIMILAR TO GROS PROTEIN (FRAGMENT)
OS Enterobacter asburiae
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Enterobacter
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-JCM6051:
RA HARADA H., ISHIKAWA H.

```

```

RL J. Gen. Appl. Microbiol. 43:355-361(1997).
DR EMBL: AB008137; BAA25206.1; -.
DR PFAM: PF00166; cpn10.1.
FT NON-TER
SQ SEQUENCE 88 AA: 9254 MW: 25045493 CRC32:

Query Match 57.8% Score 65: DB 2: Length 86:
Best Local Similarity 60.0% Pred. No. 1,140-00:
Matches 9: Conservative 3: Mismatches 3: Indels 0:

DB 72 EBYLMSRDLVLSR 86
QY 1 EBYLMSRDLVLSR 15

RESULT 13
ID 066211 PRELIMINARY: PRT: 88 AA
AC 066211
DI 01-AUG-1998 (TREMBLER: 07, Created)
DI 01-AUG-1998 (TREMBLER: 07, Last sequence update)
DE 01-MAY-1999 (TREMBLER: 10, Last annotation update)
DE SIMILAR TO GROS PROTEIN (FRAGMENT)
OS Klebsiella planticola
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Klebsiella
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-JCM 7251:
RA HARADA H., ISHIKAWA H.
RA J. Gen. Appl. Microbiol. 43:355-361(1997).
DR EMBL: AB008148; BAA25228.1; -.
DR PFAM: PF00166; cpn10.1.
FT NON-TER
SQ SEQUENCE 88 AA: 9212 MW: 25194402 CRC32:

Query Match 57.0% Score 65: DB 2: Length 86:
Best Local Similarity 60.0% Pred. No. 1,140-00:
Matches 9: Conservative 3: Mismatches 3: Indels 0: Gaps 0:

DB 72 EBYLMSRDLVLSR 86
QY 1 EBYLMSRDLVLSR 15

RESULT 14
ID 066193 PRELIMINARY: PRT: 88 AA
AC 066193
DI 01-AUG-1998 (TREMBLER: 07, Created)
DI 01-AUG-1998 (TREMBLER: 07, Last sequence update)
DE 01-MAY-1999 (TREMBLER: 10, Last annotation update)
DE SIMILAR TO GROS PROTEIN (FRAGMENT)
OS Enterobacter gergoviae
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Enterobacter
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-JCM 1234:
RA HARADA H., ISHIKAWA H.
RA J. Gen. Appl. Microbiol. 43:355-361(1997).
DR EMBL: AB008139; BAA25210.1; -.
DR PFAM: PF00166; cpn10.1.
FT NON-TER
SQ SEQUENCE 88 AA: 9239 MW: 25974000 CRC32:

Query Match 57.0% Score 65: DB 2: Length 86:
Best Local Similarity 60.0% Pred. No. 1,140-00:
Matches 9: Conservative 3: Mismatches 3: Indels 0: Gaps 0:

DB 72 EBYLMSRDLVLSR 86
QY 1 EBYLMSRDLVLSR 15

```

```

RESULT 15
ID O66219 PRELIMINARY: PRT: 88 AA
AC O66219
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DI 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
CF 01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE SIMILAR TO GROS PROTEIN (FRAGMENT).
CS Erwinia carotovora.
OC Bacteria: Proteobacteria: gamma subdivision: Enterobacteriaceae:
OC Erwinia.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAINS: IAM 12433.
RA HARADA H., ISHIKAWA H.:
RC J. Gen. Appl. Microbiol. 43:355-361(1997).
DR EMBL: AB008152; EAA25236.1;
DR PFM: PF00166; cpn10.1.
FT NON-TER 1
SQ SEQUENCE 88 AA: 9233 MW: 0ADDE5CE CRC32:

Query Match 57.0%; Score 65; DB 2; length 88;
Best Local Similarity 60.0%; Pred. NC. 1.14e-00;
Matches 9; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

DE 72 REVLMSESQILAV 86
II I I I I I I I I
QY 1 REVLSARDVAVV 15

Search completed: Mon Jun 19 16:07:23 2000
Job time : 12 secs.

```

\*\*\*\*\*



\*\*\*\*\*

Release 3.1A John F. Collins, BioComputing Research Unit,  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

Wsrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Mon Jun 19 16:05:10 2000: MasPar time 2.59 Seconds

Tabular output not generated. 82.425 Million cell updates/sec

Title: >US-09-142-524A-7

Description: (1-9) from US09142524A.pep

Sequence: 1 FIKRVSNVI 9

Scoring table: FAM 150

GAP 15

Searched: 185953 seqs, 23665106 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries  
Maximum DB seq length 9

Database: a-oneseq35  
1-oneseq

Statistics: Mean 16.765; Variance 46.25; scale 0.362

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

| Result | No. | Score | Match | Length | DB ID  | Description            | Pred. No. |
|--------|-----|-------|-------|--------|--------|------------------------|-----------|
| 1      | 54  | 100.0 | 9     | 1      | W57751 | Residues 109-117 of Cr | 1.25e-30  |

Note: Post-processor removed 44 summaries from list due to search parameters chosen.

ALIGNMENTS

RESULT 1  
ID W57751 standard: peptide 9 AA.  
AC W57751  
DI 17-SEP-1998 (first entry)  
DE Residues 109-117 of Cr  
KW Cr 117: Japanese cedar pollen antigen; allergy: immunotherapy;  
KW HLA class II molecule.  
OS Cryptomeria japonica.  
PN W09820902.AV  
PC 22-MAY-1998  
PR 12-NOV-1997: J04129.  
PR 13-NOV-1996: JP-102653.  
PA (MEIP) MEIJ: MILK PRO CO LTD.  
PI Saitaki K, Kikino K, Kume A, Sone T.  
DB WPI: 96-2376:7/25.

PT Peptides derived from Japanese cedar pollen antigens are  
PT immunotherapeutic agents - useful for allergy treatment and typing  
PT HLA class II molecules in allergy sufferers  
PS Claim 12: Page 26: 50pp: Japanese  
CC This sequence represents residues 109-117 of the Cr 117 protein and  
CC is a peptide of the invention. The peptides are derived from Japanese  
CC cedar pollen antigens, and are used as immunotherapeutic agents in the  
CC treatment of allergy. The peptides can be used for identification and  
CC typing of the particular HLA class II molecules in an allergy sufferer,  
CC and also for peptide immunotherapy of an allergy. Using these peptides  
CC the immunotherapy can be targeted more specifically to the requirements  
CC of the individual patient, allowing more effective treatment of an  
CC allergy, including those patients for whom treatment with a conventional  
CC immunotherapeutic agent is ineffective.  
SQ Sequence 9 AA:

Query Match 100.0%; Score 64; DB 1; Length 9;

Best Local Similarity 100.0%; Pred. No. 1.25e-30; Gaps 0;  
Matches 9; Conservative 0; Mismatches 0; Indels 0;

DB 1 FIKRVSNVI 9

QY 1 FIKRVSNVI 9

Search completed: Mon Jun 19 16:05:22 2000

Job time : 12 secs.

\*\*\*\*\*  
WATERBURY (TM)  
\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit,  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution Rights by Oxford Molecular Ltd

Musrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Tue Jun 20 13:39:14 2000; Waspar time 2.19 Seconds  
59.362 Million cell updates/sec

Tubular output not generated.

Title: >US-09-142-524A-7  
Description: (1-9) from US03142524A.pep  
Perfect Score: 64  
Sequence: 1 FIKRVSNVI 9

Scoring table: PAM 150  
Gap 15

Searched: 145341 seqs, 14437480 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries  
Maximum DB seq length 9

Database: a-issued  
1:5A\_COMB 2:5B\_COMB 3:6\_COMB 4:PCT\_COMB 5:backfiles1

Statistics: Mean 15.775; Variance 46.625; scale 0.338

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

| Result | Query |              |       |             |          |
|--------|-------|--------------|-------|-------------|----------|
| No.    | Score | Match length | DB ID | Description | Pred No. |

No matches found.

Search completed: Tue Jun 20 13:39:17 2000  
Job time : 3 secs.



```
No matches found.
Search completed: Mon Jun 19 16:09:17 2000
J.B. time : 23 secs.
```

.....  
MPSRELA (TM)  
.....

Release 3.1A John F. Collins, BioComputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd  
Mpsrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Mon Jun 19 16:04:38 2000 MasPar time 4.43 Seconds  
95.841 Million cell updates/sec  
Tabular output not generated.

Title: >US-09-142-524A-7  
Description: (179) from USC9142524A.pep  
Perfect Score: 54  
Sequence: 1 FKKVSNV: 9  
Scoring table: PAM 150  
Gap 15

Searched: 142080 seqs, 47172406 residues  
Fast-processing: Minimum Match 0%  
Listing first 45 summaries  
Maximum DB seq length 9

Database: pir62  
1:pir1 2:pir2 3:pir3 4:pir4  
Statistics: Mean 23.798; Variance 28.919; scale 0.823

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES  
Result %  
Query  
No. Score Match Length DB ID Description Pred. No.  
.....

No matches found.  
Search completed: Mon Jun 19 16:04:53 2000  
Job time: 15 secs.

\*\*\*\*\*  
[WATER] (TM)  
\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsearch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Mon Jun 19 16:03:40 2000: MasPar time 3.37 Seconds  
81.234 Million cell updates/sec

Tabular output not generated.

Title: >US-09-142-524A-7  
Description: (1-9) from USC9142524A.pep  
Perfect score: 64  
Sequence: 1 FIKRVSNVI 9

Scoring table: PAM 150  
Gap 15

Searched: 83857 seqs, 30454973 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries  
Maximum DB seq length 9

Database: swiss-prot38  
1:swissprot

Statistics: Mean 24.343: Variance 25.165: scale 0.967

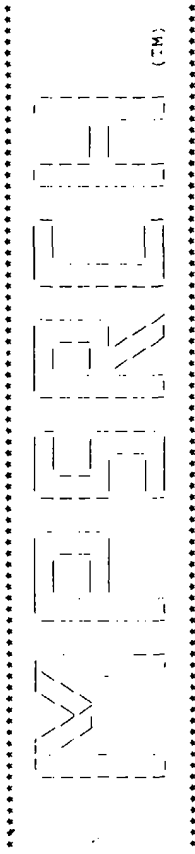
Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

| Result | Query |       |        |    |             |
|--------|-------|-------|--------|----|-------------|
| No.    | Score | Match | Length | ID | Description |
|        |       |       |        |    | Prod. No.   |

No matches found.

Search completed: Mon Jun 19 16:03:47 2000  
Job time : 7 secs.



Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

Msrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Mon Jun 19 16:04:03 2000; MasPar time 8.30 Seconds  
75.200 Million cell updates/sec

Tabular output not generated.

Title: >US-09-142-524A-7  
Description: (1-9) from US09142524A.peg  
Perfect Score: 64  
Sequence: 1 FKRVSNV: 9

Scoring table: PAM 150  
Gap 15

Searched: 225878 seqs, 69334122 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries  
Maximum DB seq length 9

Database: sptrembl2  
1:sp-archaea 2:sp-bacteria 3:sp-fungi 4:sp-human  
5:sp-invertebrate 6:sp-mammal 7:sp-mhc 8:sp-organelle  
9:sp-phage 10:sp-plant 11:sp-rodent 12:sp-unclassified  
13:sp-vertebrate 14:sp-virus

Statistics: Mean 23.348; Variance 25.366; scale 0.923

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

| Result            | Query | Match | length | DB ID | Description | Pred. No. |
|-------------------|-------|-------|--------|-------|-------------|-----------|
| No matches found. |       |       |        |       |             |           |

Search completed: Mon Jun 19 16:04:22 2000  
Job time : 19 secs.

.....  
WOREL  
..... (TW)  
.....

Release 3.1A John F. Collins, BioComputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd.  
Msearch\_pp proteins - protein database search, using Smith-Waterman algorithm  
Run on: Mon Jun 19 16:10:20 2000; Maspar time 3.31 Seconds  
Tabular output not generated. 107.475 Million cell updates/sec

Title: >US-09-142-524A-9  
Description: (1-15) from US09142524A.pep  
Perfect Score: 100  
Sequence: 1 KSMKVTVAFNQFGPN 15  
Scoring table: PAM 150  
Gap 15  
Searched: 166963 seqs, 21686106 residues  
Post-processing: Minimum Match 0%  
Listing first 45 summaries  
Database: a-seqseq35  
igeneseq  
Statistics: Mean 18.458; Variance 56.349; scale 0.328

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | ID  | Description             | Pred. No. |
|------------|-------|-------------|-----|-------------------------|-----------|
| 1          | 106   | 100.0       | 15  | Japanese cedar pollen   | 8.38e-05  |
| 2          | 106   | 100.0       | 15  | Residues 211-225 of Cr  | 8.38e-05  |
| 3          | 106   | 100.0       | 17  | Sugi allergen protein   | 8.38e-05  |
| 4          | 106   | 100.0       | 20  | T-cell epitope peptide  | 8.38e-05  |
| 5          | 106   | 100.0       | 20  | W42144                  | 8.38e-05  |
| 6          | 106   | 100.0       | 20  | R82512                  | 8.38e-05  |
| 7          | 106   | 100.0       | 26  | Cry j 1 Japanese cedar  | 8.38e-05  |
| 8          | 106   | 100.0       | 26  | R45592                  | 8.38e-05  |
| 9          | 106   | 100.0       | 26  | Cry j 1 pollen allergen | 8.38e-05  |
| 10         | 106   | 100.0       | 26  | R45594                  | 8.38e-05  |
| 11         | 106   | 100.0       | 28  | Cry j 1 pollen allergen | 8.38e-05  |
| 12         | 106   | 100.0       | 28  | R45590                  | 8.38e-05  |
| 13         | 106   | 100.0       | 28  | Cry j 1 pollen allergen | 8.38e-05  |
| 14         | 106   | 100.0       | 30  | R45589                  | 8.38e-05  |
| 15         | 106   | 100.0       | 30  | Cry j 1 pollen allergen | 8.38e-05  |
| 16         | 106   | 100.0       | 30  | R45588                  | 8.38e-05  |
| 17         | 106   | 100.0       | 36  | T-cell epitope peptide  | 8.38e-05  |
| 18         | 106   | 100.0       | 36  | W44587                  | 8.38e-05  |
| 19         | 106   | 100.0       | 353 | Cry j 1 pollen allergen | 8.38e-05  |
| 20         | 106   | 100.0       | 353 | R75388                  | 8.38e-05  |
| 21         | 106   | 100.0       | 354 | Japanese cedar pollen   | 8.38e-05  |
| 22         | 106   | 100.0       | 354 | Japanese cypress polle  | 8.38e-05  |
| 23         | 106   | 100.0       | 367 | Chamaecyparis obtusa p  | 8.38e-05  |
| 24         | 106   | 100.0       | 374 | Jun s 1                 | 8.38e-05  |
| 25         | 106   | 100.0       | 374 | Cry j 1 pollen allergen | 8.38e-05  |
| 26         | 106   | 100.0       | 374 | Japanese cedar pollen   | 8.38e-05  |

24 106 100.0 374 1 R31937  
25 106 100.0 374 1 R2490  
26 106 100.0 375 1 W4345  
27 97 91.5 370 1 R45578  
28 95 89.6 80 1 W2369  
29 95 89.6 105 1 W2370  
30 95 89.6 134 1 W2371  
31 94 88.7 13 1 Y01756  
32 94 88.7 81 1 W60358  
33 93 87.7 22 1 R2178  
34 85 85.2 47 1 W60350  
35 85 83.2 47 1 W60351  
36 83 78.3 47 1 W60353  
37 83 78.3 47 1 W60356  
38 82 77.4 11 1 Y01745  
39 82 77.4 47 1 W60357  
40 82 77.4 47 1 W60352  
41 78 73.6 401 1 W57846  
42 68 64.2 20 1 R62511  
43 68 64.2 20 1 W42143  
44 68 64.2 20 1 R45562  
45 64 60.4 9 1 W57759

ALIGNMENTS

RESULT 1  
ID R83293 standard; peptide: 15 AA.  
AC R83293;  
DT 12-MAR-1996 (first entry)  
DE Japanese cedar pollen allergen Cry j 1 derived T-cell epitope peptide  
KW Japanese cedar; pollen allergen; Cry j 1 T-cell epitope peptides.  
KW prevention; treatment; cryptomeria pollenosis.  
OS Cryptomeria japonica.  
PN J07118295-A.  
PD 09-MAY-1995.  
PF 20-OCT-1993; 262626.  
PR 20-OCT-1993; JP-262626.  
PA (MEIP ) MEIJI MILK PROD CO. LTD.  
DR WPI; 95-203834/27.  
PT New cryptomeria pollen allergen T-cell epitope peptide - used for prevention, treatment and investigation of Japanese cedar pollenosis  
PS Claim 5; page 2; 8pp; Japanese  
CC R73388 is the Japanese cedar pollen allergen Cry j 1 from which the T-cell epitope peptides R83293-R83295 were derived. The peptides can be used for the prevention and treatment of cryptomeria pollenosis, and also for the investigation of pollenosis.  
SQ Sequence 15 AA:  
Query Match 100.0%; Score 106; DB 1; Length 15.  
Best Local Similarity 100.0%; Pred. No. 8.38e-05.  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0.

Ddb 1 KSMKVTVAFNQFGPN 15  
|||||  
QY 1 KSMKVTVAFNQFGPN 15

RESULT 2  
ID W57755 standard; peptide: 15 AA.  
AC W57755;  
DT 17-SEP-1998 (first entry)  
DE Residues 211-225 of Cry j 1.  
KW Cry j 1; Japanese cedar pollen antigen; allergy; immunotherapy.  
OS HLA class II molecule.  
PN W08820902-A1.  
PD 22-MAY-1998.  
PF 12-NOV-1997; J04129.  
PR 13-NOV-1996; JP-302053.  
PA (MEIP ) MEIJI MILK PROD CO. LTD.  
PI Daiichi K. Kino K, Kume A, Sone T;  
DR WPI; 93-297617/26.

7-cell epitope peptide portion of Japanese cypress pollen antigens  
Chao1 and Chao2 - used for diagnosis and treatment of Spring tree  
pollen disease  
Claim 1: Page 27-28; 71pp; Japanese  
The present sequence represents a 7-cell epitope peptide from Japanese  
cypress pollen antigen Chao1. The present invention describes peptides  
which correspond to the T-cell epitope sites on Japanese cypress pollen  
antigens Chao1 and Chao2. The peptides can be used as a reagent for the  
diagnosis of allergy to Japanese cypress pollen, and as an antigen in  
the treatment and prevention of Spring tree pollen disease in which the  
pollinosis involves reactivity to Japanese cypress pollen.  
Sequence 20 AA;  
SQ

```

Query Match      100.0%   Score 106   DB i, Length 20
Best Local Similarity 100.0%   Pred. No. 8,38e-05
Matches         15: Conservative   0: Mismatches   0: Indels   0: Gaps

```

|    | Matches | Conservative      | X:Small | Y:Large |
|----|---------|-------------------|---------|---------|
| Db | 1       | KSMKVVAFNQFGPN 15 |         |         |
|    |         | ! ! ! ! ! ! ! !   |         |         |
| QY | 1       | KSMKVVAFNQFGPN 15 |         |         |

RC 802512:  
DT 15-APR-1996 (first entry)  
DE Cry 3 Japanese Cedar pollen allergen peptide fragment (321-22)  
KW Cry 3 Japanese cedar pollen allergen, modified drug production

Cryptomeria japonica allergen; modified drug production.  
KW Japanese cedar pollen allergen;  
KW Cryptomeria japonica  
CS Cryptomeria japonica  
PN W09S7786-A

PF 06-APR-1993: US-249.  
PR 08-APR-1994: US-226248.  
PR 06-DEC-1994: US-35025.  
PA (IMMUN-) IMMUNOLOGIC PHARM CORP.

PT Modified *Cryptomeria japonica* (Cry J) peptides(s) + usutu. for  
PT treating allergy to Japanese cedar pollen allergen or  
PT immunologically cross reactive allergens  
PS Discussed: Figure 2, 6000 English

CC possess certain characteristics which render them particularly  
CC suitable for drug product formulation. Peptide fragments of Gly 1-  
CC modified and unmodified, are given in KR249-KR525. This peptide  
CC fragment corresponds to amino acids 211-230 of the allergen mature  
CC protein.  
CC Sequence 20 AA.

Db: 1 KSMKVTVAENQFGPN 15  
Matches: 15: Conservative 0: N:smatches 0: Injels 0: caps 0: caps

| RESULT  | 5   |
|---------|---|
| ID      | R45563 standard; Protein: 20 AA.                          |
| R45563: |   |
| AC      |   |
| DE      | 13-JUL-1994 (first entry)                                 |
| DE      | Cry 3; polices: allergen: peptide CJI-22.                 |
| KW      | Japanese Cedar; detection: allergy; treatment: diagnosis: |
| KW      | I ced.; epitope: sensitivity.                             |
| OS      | Cryptomeria japonica.                                     |
| PN      | W094C1560.A.  |

20-JAN-1994.  
 PF 15-JAN-1993: U00139.  
 PR 10-JUL-1992: WO-005661.  
 PA 01-SEP-1992: US-938990.  
 PA (IMMU-) IMMUNOLOGIC PHARM CORP.  
 PI Bond JF, Garman RD, Griffith IJ, Kuo M, Pollock J.  
 DR WPI: 94-035066/04.  
 PT Antigens derived from Japanese cedar pollen allergen Cry j 1.  
 PI contain at least two T cell epitope(s), used to treat or diagnose  
 PT allergy.  
 PS Claim 76: Fig 18: 137pp: English.  
 CC The sequence is that of an isolated peptide of the Japanese cedar  
 CC pollen allergen Cry j 1 (amino acids 211-230). The peptide, CJI-43.8,  
 CC can be used for the treatment and diagnosis of allergies associated  
 CC with Japanese cedar pollen. It has enhanced therapeutic properties  
 CC but reduced side effects compared to naturally occurring allergens.  
 SQ Sequence 26 AA:

Query Match 100.0%; Score 106; DB 1; Length 26;  
 Best Local Similarity 100.0%; Pred. No. 8.38e-05;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 1 KSMKVTVAENQFGPN 15  
 QY 1 KSMKVTVAENQFGPN 15

RESULT 7  
 ID R45592 standard; Protein: 26 AA.  
 AC R45592:  
 DE 13-JUL-1994 (first entry)  
 DE Cry j 1 pollen allergen peptide CJI-43.8.  
 KW Japanese cedar; detection: allergy; treatment: diagnosis;  
 KW T cell epitope; sensitivity.  
 OS Cryptomeria japonica.  
 PN W09401560-A.  
 PD 20-JAN-1994.  
 PF 15-JAN-1993: U00139.  
 PR 10-JUL-1992: WO-005661.  
 PR 01-SEP-1992: US-938990.  
 PA (IMMU-) IMMUNOLOGIC PHARM CORP.  
 PI Bond JF, Garman RD, Griffith IJ, Kuo M, Pollock J.  
 DR WPI: 94-035066/04.  
 PT Antigens derived from Japanese cedar pollen allergen Cry j 1.  
 PI contain at least two T cell epitope(s), used to treat or diagnose  
 PT allergy.  
 PS Claim 76: Fig 18: 137pp: English.  
 CC The sequence is that of an isolated peptide of the Japanese cedar  
 CC pollen allergen Cry j 1. The peptide, CJI-43.8, can be used for  
 CC the treatment and diagnosis of allergies associated with Japanese  
 CC cedar pollen. It has enhanced therapeutic properties but reduced  
 CC side effects compared to naturally occurring allergens.  
 SQ Sequence 26 AA:

Query Match 100.0%; Score 106; DB 1; Length 26;  
 Best Local Similarity 100.0%; Pred. No. 8.38e-05;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 1 KSMKVTVAENQFGPN 15  
 QY 1 KSMKVTVAENQFGPN 15

RESULT 8  
 ID R45591 standard; Protein: 26 AA.  
 AC R45591:  
 DE 13-JUL-1994 (first entry)  
 DE Cry j 1 pollen allergen peptide CJI-43.9.  
 KW Japanese cedar; detection: allergy; treatment: diagnosis;  
 KW T cell epitope; sensitivity.  
 OS Cryptomeria japonica.  
 PN W09401560-A.  
 PD 20-JAN-1994.

15-JAN-1993: U00139.  
 PR 10-JUL-1992: WO-005661.  
 PR 01-SEP-1992: US-938990.  
 PA (IMMU-) IMMUNOLOGIC PHARM CORP.  
 PI Bond JF, Garman RD, Griffith IJ, Kuo M, Pollock J.  
 DR WPI: 94-035066/04.  
 PT Antigens derived from Japanese cedar pollen allergen Cry j 1.  
 PI contain at least two T cell epitope(s), used to treat or diagnose  
 PT allergy.  
 PS Claim 76: Fig 18: 137pp: English.  
 CC The sequence is that of an isolated peptide of the Japanese cedar  
 CC pollen allergen Cry j 1. The peptide, CJI-43.9, can be used for  
 CC the treatment and diagnosis of allergies associated with Japanese  
 CC cedar pollen. It has enhanced therapeutic properties but reduced  
 CC side effects compared to naturally occurring allergens.  
 SQ Sequence 26 AA:

Query Match 100.0%; Score 106; DB 1; Length 26;  
 Best Local Similarity 100.0%; Pred. No. 8.38e-05;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 1 KSMKVTVAENQFGPN 15  
 QY 1 KSMKVTVAENQFGPN 15

RESULT 9  
 ID R45594 standard; Protein: 26 AA.  
 AC R45594:  
 DE 13-JUL-1994 (first entry)  
 DE Cry j 1 pollen allergen peptide CJI-43.12.  
 KW Japanese cedar; detection: allergy; treatment: diagnosis;  
 KW T cell epitope; sensitivity.  
 OS Cryptomeria japonica.  
 PN W09401560-A.  
 PD 20-JAN-1994.  
 PF 15-JAN-1993: U00139.  
 PR 10-JUL-1992: WO-005661.  
 PR 01-SEP-1992: US-938990.  
 PA (IMMU-) IMMUNOLOGIC PHARM CORP.  
 PI Bond JF, Garman RD, Griffith IJ, Kuo M, Pollock J.  
 DR WPI: 94-035066/04.  
 PT Antigens derived from Japanese cedar pollen allergen Cry j 1.  
 PI contain at least two T cell epitope(s), used to treat or diagnose  
 PT allergy.  
 PS Claim 76: Fig 18: 137pp: English.  
 CC The sequence is that of an isolated peptide of the Japanese cedar  
 CC pollen allergen Cry j 1. The peptide, CJI-43.12, can be used for  
 CC the treatment and diagnosis of allergies associated with Japanese  
 CC cedar pollen. It has enhanced therapeutic properties but reduced  
 CC side effects compared to naturally occurring allergens.  
 SQ Sequence 26 AA:

Query Match 100.0%; Score 106; DB 1; Length 26;  
 Best Local Similarity 100.0%; Pred. No. 8.38e-05;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 1 KSMKVTVAENQFGPN 15  
 QY 1 KSMKVTVAENQFGPN 15

RESULT 10  
 ID R45590 standard; Protein: 28 AA.  
 AC R45590:  
 DE 13-JUL-1994 (first entry)  
 DE Cry j 1 pollen allergen peptide CJI-43.8.  
 KW Japanese cedar; detection: allergy; treatment: diagnosis;  
 KW T cell epitope; sensitivity.  
 OS Cryptomeria japonica.  
 PN W09401560-A.  
 PD 20-JAN-1994.  
 PF 15-JAN-1993: U00139.

PR 10-JUL-1992: WO-005661.  
 PA (IMMU-) IMMUNOLOGIC PHARM CORP.  
 PI Bond JF, Garman RD, Griffith LJ, Kuo M, Pollock J;  
 DR WPI: 94-035066/04.  
 DE Anticidens derived from Japanese cedar pollen allergen Cry j 1.  
 PI contain at least two T cell epitope(s), used to treat or diagnose  
 PT allergy.  
 PS Claim 76: Fig 18: 137pp: English.  
 CC The sequence is that of an isolated peptide of the Japanese cedar  
 CC pollen allergen Cry j 1. The peptide, CJI-43.8, can be used for  
 CC the treatment and diagnosis of allergies associated with Japanese  
 CC cedar pollen. It has enhanced therapeutic properties but reduced  
 CC side effects compared to naturally occurring allergens.  
 SQ Sequence 28 AA.

Query Match 100.0% Score 106; DB 1; Length 28;  
 Best Local Similarity 100.0%; Pred. No. 8.38e-05;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DE 1 KSMKVTVAFNQFGPN 15  
 ||||| ||||| |||||  
 QY 1 KSMKVTVAFNQFGPN 15

RESULT 11  
 ID R45593 standard: Protein: 28 AA.

AC R45593:  
 DT 13-JUL-1994 (first entry)  
 DE Cry j 1 pollen allergen peptide CJI-43.11.  
 KW Japanese cedar; detection: allergy; treatment: diagnosis;  
 KW T cell epitope; sensitivity.  
 OS Cryptomeria japonica.  
 PN WO9401560-A.  
 PD 20-JAN-1994.  
 PF 15-JAN-1993; U00139.  
 PR 10-JUL-1992: WO-005661.  
 PA (IMMU-) IMMUNOLOGIC PHARM CORP.  
 PI Bond JF, Garman RD, Griffith LJ, Kuo M, Pollock J;  
 DR WPI: 94-035066/04.  
 DE Anticidens derived from Japanese cedar pollen allergen Cry j 1.  
 PI contain at least two T cell epitope(s), used to treat or diagnose  
 PT allergy.  
 PS Claim 76: Fig 18: 137pp: English.  
 CC The sequence is that of an isolated peptide of the Japanese cedar  
 CC pollen allergen Cry j 1. The peptide, CJI-43.11, can be used for  
 CC the treatment and diagnosis of allergies associated with Japanese  
 CC cedar pollen. It has enhanced therapeutic properties but reduced  
 CC side effects compared to naturally occurring allergens.  
 SQ Sequence 28 AA.

Query Match 100.0% Score 106; DB 1; Length 28;  
 Best Local Similarity 100.0%; Pred. No. 8.38e-05;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DE 1 KSMKVTVAFNQFGPN 15  
 ||||| ||||| |||||  
 QY 1 KSMKVTVAFNQFGPN 15

RESULT 12  
 ID R45589 standard: Protein: 28 AA.

AC R45589:  
 DT 13-JUL-1994 (first entry)  
 DE Cry j 1 pollen allergen peptide CJI-43.7.  
 KW Japanese cedar; detection: allergy; treatment: diagnosis;  
 KW T cell epitope; sensitivity.  
 OS Cryptomeria japonica.  
 PN WO9401560-A.  
 PD 20-JAN-1994.  
 PF 15-JAN-1993; U00139.  
 PR 10-JUL-1992: WO-005661.

PR 01-SEP-1992: US-938990.  
 PA (IMMU-) IMMUNOLOGIC PHARM CORP.  
 PI Bond JF, Garman RD, Griffith LJ, Kuo M, Pollock J;  
 DR WPI: 94-035066/04.  
 DE Anticidens derived from Japanese cedar pollen allergen Cry j 1.  
 PI contain at least two T cell epitope(s), used to treat or diagnose  
 PT allergy.  
 PS Claim 76: Fig 18: 137pp: English.  
 CC The sequence is that of an isolated peptide of the Japanese cedar  
 CC pollen allergen Cry j 1. The peptide, CJI-43.7, can be used for  
 CC the treatment and diagnosis of allergies associated with Japanese  
 CC cedar pollen. It has enhanced therapeutic properties but reduced  
 CC side effects compared to naturally occurring allergens.  
 SQ Sequence 28 AA.

Query Match 100.0% Score 106; DB 1; Length 28;  
 Best Local Similarity 100.0%; Pred. No. 8.38e-05;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DE 1 KSMKVTVAFNQFGPN 15  
 ||||| ||||| |||||  
 QY 1 KSMKVTVAFNQFGPN 15

RESULT 13  
 ID R45587 standard: Protein: 30 AA.

AC R45587:  
 DT 13-JUL-1994 (first entry)  
 DE Cry j 1 pollen allergen peptide CJI-43.1.  
 KW Japanese cedar; detection: allergy; treatment: diagnosis;  
 KW T cell epitope; sensitivity.  
 OS Cryptomeria japonica.  
 PN WO9401560-A.  
 PD 20-JAN-1994.  
 PF 15-JAN-1993; U00139.  
 PR 10-JUL-1992: WO-005661.  
 PA (IMMU-) IMMUNOLOGIC PHARM CORP.  
 PI Bond JF, Garman RD, Griffith LJ, Kuo M, Pollock J;  
 DR WPI: 94-035066/04.  
 DE Anticidens derived from Japanese cedar pollen allergen Cry j 1.  
 PI contain at least two T cell epitope(s), used to treat or diagnose  
 PT allergy.  
 PS Claim 76: Fig 18: 137pp: English.  
 CC The sequence is that of an isolated peptide of the Japanese cedar  
 CC pollen allergen Cry j 1. The peptide, CJI-43.1, can be used for  
 CC the treatment and diagnosis of allergies associated with Japanese  
 CC cedar pollen. It has enhanced therapeutic properties but reduced  
 CC side effects compared to naturally occurring allergens.  
 SQ Sequence 30 AA.

Query Match 100.0% Score 106; DB 1; Length 30;  
 Best Local Similarity 100.0%; Pred. No. 8.38e-05;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DE 1 KSMKVTVAFNQFGPN 15  
 ||||| ||||| |||||  
 QY 1 KSMKVTVAFNQFGPN 15

RESULT 14  
 ID R45588 standard: Protein: 30 AA.

AC R45588:  
 DT 13-JUL-1994 (first entry)  
 DE Cry j 1 pollen allergen peptide CJI-43.6.  
 KW Japanese cedar; detection: allergy; treatment: diagnosis;  
 KW T cell epitope; sensitivity.  
 OS Cryptomeria japonica.  
 PN WO9401560-A.  
 PD 20-JAN-1994.  
 PF 15-JAN-1993; U00139.  
 PR 10-JUL-1992: WO-005661.  
 PR 01-SEP-1992: US-938990.



PA (IMMUNO) IMMUNOLOGIC PHARM CORP.  
 PI Bond JF, Garman RD, Griffith LJ, Kuo M, Collock J.  
 PR WPI: 94-035065/04.  
 P: Antigens derived from Japanese cedar pollen allergen Cry 1-1  
 R: contain at least two T cell epitope(s), used to treat or diagnose  
 E: allergy.  
 FS Claim 76; Fig 18; 137pp; English.  
 CO The sequence is that of an isolated peptide of the Japanese cedar  
 Q: pollen allergen Cry 1. The peptide, Cyl-43.6, can be used for  
 C: the treatment and diagnosis of allergies associated with Japanese  
 Q: cedar pollen. It has enhanced therapeutic properties but reduced  
 Q: side effects compared to naturally occurring allergens.  
 SQ Sequence 30 AA:

Query Match 100.0% Score 100; 31; Length 30;  
 Best Local Similarity 100.0% Pred. No. 8, 18e-05;  
 Matches 15; Conservative 0; Mismatches 0; Gaps 0;


DA 1 KSMKVTVAENQFGPN 15  
 QY 1 KSMK VARNQFGPN 15

RESULT 15  
 ID W44587 standard: peptide: 30 AA.  
 AC W44587.  
 DT 01-MAY-1998 (first entry)  
 EE T-cell epitope peptide #6 of sugi pollen antigen.  
 KW T-cell epitope: sugi pollen antigen; sugi pollinosis.  
 OS Synthesis.  
 PS CRYPTOMERIA japonica  
 PN J10007760-A.  
 PD 13-JAN-1998.  
 PI 24-JUN-1998; 163287.  
 PK 24-JUN-1998; JP-163287.  
 RA (DAIIC) DAIICHEM IND LTD.  
 SA (MELB) MELB SEIKA KATSHA LTD.  
 SW WPI: 98-035690/15.  
 PT T cell epitope peptide of sugi pollen antigen - useful in the  
 E: treatment of sugi pollinosis  
 PS Claim 1; Page 4; 14pp; Japanese.  
 Q: T-cell epitope peptides W44582-88 and their derivatives react with  
 C: sugi pollinosis patient peripheral blood T lymphocytes. A composition  
 Q: prepared by combining at least 2 of the above peptides and/or their  
 Q: derivatives is used for the prevention and treatment of sugi pollinosis.  
 SQ Sequence 30 AA:

Query Match 100.0% Score 100; 31; Length 30;  
 Best Local Similarity 100.0% Pred. No. 8, 18e-05;  
 Matches 15; Conservative 0; Mismatches 0; Gaps 0;

DA 11 KSMKVTVAENQFGPN 25  
 QY 1 KSMKVTVAENQFGPN 15

Search completed: Mon Jun 19 16:10:31 2000  
 Job time: 11 sec



王 國 強 民 強 國

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

```

MSearch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Tue Jun 20 13:35:56 2000: Maspas time 2.65 Seconds
81.869 Million cell updates/sec
Isular output not generated.

```

```

Title:
Description:
Perfect Score:
Sequence:
Scoring table:

```

Searched: 145341 seqs. 14437480 residues

Post-processing: Minimum Match 08  
Listing first 45

```
Database: a..Issued
1: 5A_COMB 2: 5B_COMB 3: 5_COMB 4: PCT_COMB 5: backfiles1
```

Statistics: Mean: 7.411; Variance 53.947; scale 0.323

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Query |        | DB | ID         | Description  | Pred. No. |
|------------|-------|-------|--------|----|------------|--------------|-----------|
|            |       | Match | Length |    |            |              |           |
| 1          | 57    | 53.8  | 397    | 1  | US-08-175- | Sequence 76, | 2.10e-01  |
| 2          | 57    | 53.8  | 397    | 1  | US-08-200- | Sequence 76, | 2.10e-01  |
| 3          | 57    | 53.8  | 397    | 1  | US-08-230- | Sequence 76, | 2.10e-01  |
| 4          | 55    | 51.9  | 391    | 1  | US-08-290- | Sequence 59, | 3.43e-01  |
| 5          | 55    | 51.9  | 391    | 1  | US-08-290- | Sequence 59, | 3.43e-01  |
| 6          | 55    | 51.9  | 391    | 1  | US-08-175- | Sequence 59, | 3.43e-01  |
| 7          | 55    | 51.9  | 398    | 1  | US-08-200- | Sequence 74, | 3.43e-01  |
| 8          | 55    | 51.9  | 398    | 1  | US-08-200- | Sequence 74, | 3.43e-01  |
| 9          | 55    | 51.9  | 398    | 1  | US-08-175- | Sequence 74, | 3.43e-01  |
| 10         | 54    | 50.9  | 387    | 1  | US-08-290- | Sequence 72, | 4.37e-01  |
| 11         | 54    | 50.9  | 387    | 1  | US-08-290- | Sequence 72, | 4.37e-01  |
| 12         | 54    | 50.9  | 387    | 2  | US-08-175- | Sequence 72, | 4.37e-01  |
| 13         | 53    | 50.0  | 383    | 1  | US-08-175- | Sequence 78, | 4.57e-01  |
| 14         | 53    | 50.0  | 383    | 1  | US-08-230- | Sequence 78, | 4.57e-01  |
| 15         | 53    | 50.0  | 383    | 1  | US-08-290- | Sequence 78, | 4.57e-01  |
| 16         | 52    | 49.1  | 333    | 2  | US-08-436- | Sequence 4,  | 7.09e-01  |
| 17         | 52    | 49.1  | 333    | 4  | PCT-US95-0 | Sequence 4,  | 7.09e-01  |
| 18         | 52    | 49.1  | 333    | 4  | PCT-US95-0 | Sequence 2,  | 7.09e-01  |
| 19         | 52    | 49.1  | 333    | 4  | US-08-436- | Sequence 4,  | 7.09e-01  |
| 20         | 52    | 49.1  | 336    | 4  | PCT-US95-0 | Sequence 2,  | 7.09e-01  |
| 21         | 52    | 49.1  | 336    | 1  | US-08-436- | Sequence 2,  | 7.09e-01  |
| 22         | 52    | 49.1  | 336    | 2  | US-08-436- | Sequence 2,  | 7.09e-01  |
| 23         | 52    | 49.1  | 442    | 2  | US-08-687- | Sequence 6,  | 7.09e-01  |

DB 256 KSMKLVIAVFNQFGPN 15  
LENGTH: 397 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE 397 AA: 42947 MW: 781425 CN:  
Query Match 53.8% Score 57: DB 1: Length 397:  
Best Local Similarity 60.0% Pred. No. 2.10e-01:  
Matches 9: Conservative 1: Mismatches 5: Indels 0: Gaps 0:

DB 256 KSMKLVIAVFNQFGPN 15  
LENGTH: 397 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE 397 AA: 42947 MW: 781425 CN:  
Query Match 53.8% Score 57: DB 1: Length 397:  
Best Local Similarity 60.0% Pred. No. 2.10e-01:  
Matches 9: Conservative 1: Mismatches 5: Indels 0: Gaps 0:

DB 256 KSMKLVIAVFNQFGPN 15  
LENGTH: 397 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE 397 AA: 42947 MW: 781425 CN:  
Query Match 53.8% Score 57: DB 1: Length 397:  
Best Local Similarity 60.0% Pred. No. 2.10e-01:  
Matches 9: Conservative 1: Mismatches 5: Indels 0: Gaps 0:

DB 256 KSMKLVIAVFNQFGPN 15  
LENGTH: 397 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE 397 AA: 42947 MW: 781425 CN:  
Query Match 53.8% Score 57: DB 1: Length 397:  
Best Local Similarity 60.0% Pred. No. 2.10e-01:  
Matches 9: Conservative 1: Mismatches 5: Indels 0: Gaps 0:

DB 256 KSMKLVIAVFNQFGPN 15  
LENGTH: 397 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE 397 AA: 42947 MW: 781425 CN:  
Query Match 53.8% Score 57: DB 1: Length 397:  
Best Local Similarity 60.0% Pred. No. 2.10e-01:  
Matches 9: Conservative 1: Mismatches 5: Indels 0: Gaps 0:

DB 256 KSMKLVIAVFNQFGPN 15  
LENGTH: 397 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE 397 AA: 42947 MW: 781425 CN:  
Query Match 53.8% Score 57: DB 1: Length 397:  
Best Local Similarity 60.0% Pred. No. 2.10e-01:  
Matches 9: Conservative 1: Mismatches 5: Indels 0: Gaps 0:



CC SOFTWARE: PatentIn Release #1.0, Version #1.25  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/175,069A  
CC FILING DATE: December 29, 1993  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 07/529,951  
CC FILING DATE: May 29, 1990  
CC APPLICATION NUMBER: US 07/325,365  
CC FILING DATE: March 17, 1989  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Amy E. Mandragouras  
CC REGISTRATION NUMBER: 36,207  
CC REFERENCE/DOCKET NUMBER: IMI-018CN  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (617)227-7400  
CC TELEFAX: (617)227-5941  
CC INFORMATION FOR SEQ ID NO: 59:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 391 amino acids  
CC TYPE: amino acid  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: protein  
CC SEQUENCE 391 AA: 42864 MW: 754045 CN:  
CC  
CC Query Match 51.9% Score 55: DB 1: Length 391:  
CC Best Local Similarity 66.7% Pred. NO. 3,430-01:  
CC Matches 8: Conservative 1: Mismatches 3: Indels 0: Gaps 0:  
CC  
CC Db 250 KGMALAIWAFNMF 261  
CC 1 1 1 1 1  
CC 1 KSMKVTVAFNMF 12  
CC  
CC RESULT 7  
CC US-08-290-448A-74 STANDARD: PRI: 398 AA:  
CC XXXXXX  
CC  
CC Sequence 74, Application US/08290448A  
CC Patent No. 5676554  
CC GENERAL INFORMATION:  
CC APPLICANT: Rogers, Bruce  
CC APPLICANT: Klapper, David G.  
CC APPLICANT: Rafnar, Thorunn  
CC APPLICANT: Kuo, Mei-chang  
CC TITLE OF INVENTION: Allergenic Proteins From Radweed and Uses  
CC NUMBER OF SEQUENCES: 93  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: LAHIVE & COCKFIELD  
CC STREET: 60 State Street, suite 510  
CC CITY: Boston  
CC STATE: Massachusetts  
CC COUNTRY: USA  
CC ZIP: 02109-1875  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: PatentIn Release #1.0, Version #1.25  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/290,448A  
CC FILING DATE: August 15, 1994  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 07/529,951  
CC FILING DATE: May 29, 1990  
CC APPLICATION NUMBER: US 07/325,365  
CC FILING DATE: March 17, 1989  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Amy E. Mandragouras  
CC REGISTRATION NUMBER: 36,207  
CC REFERENCE/DOCKET NUMBER: IMI-018CN  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (617)227-7400  
CC TELEFAX: (617)227-5941  
CC INFORMATION FOR SEQ ID NO: 59:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 391 amino acids  
CC TYPE: amino acid  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: protein

CC REGISTRATION NUMBER: 36,207  
CC REFERENCE/DOCKET NUMBER: IMI-018CN  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (617)227-7400  
CC TELEFAX: (617)227-5941  
CC INFORMATION FOR SEQ ID NO: 74:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 398 amino acids  
CC TYPE: amino acid  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: protein  
CC SEQUENCE 398 AA: 43564 MW: 779000 CN:  
CC  
CC Query Match 51.9% Score 55: DB 1: Length 398:  
CC Best Local Similarity 66.7% Pred. NO. 3,440-01:  
CC Matches 8: Conservative 1: Mismatches 3: Indels 0: Gaps 0:  
CC  
CC Db 257 KGMALAIWAFNMF 268  
CC 1 1 1 1 1 1  
CC 1 KSMKVTVAFNMF 12  
CC  
CC RESULT 8  
CC US-08-290-448A-74 STANDARD: PRI: 498 AA:  
CC XXXXXX  
CC  
CC Sequence 74, Application US/08290448A  
CC Patent No. 5698204  
CC GENERAL INFORMATION:  
CC APPLICANT: Rogers, Bruce  
CC APPLICANT: Klapper, David G.  
CC APPLICANT: Rafnar, Thorunn  
CC APPLICANT: Kuo, Mei-chang  
CC TITLE OF INVENTION: Allergenic Proteins From Radweed and Uses  
CC NUMBER OF SEQUENCES: 93  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: LAHIVE & COCKFIELD  
CC STREET: 60 State Street, suite 510  
CC CITY: Boston  
CC STATE: Massachusetts  
CC COUNTRY: USA  
CC ZIP: 02109-1875  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: PatentIn Release #1.0, Version #1.25  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/290,448A  
CC FILING DATE: August 15, 1994  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 07/529,951  
CC FILING DATE: May 29, 1990  
CC APPLICATION NUMBER: US 07/325,365  
CC FILING DATE: March 17, 1989  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Amy E. Mandragouras  
CC REGISTRATION NUMBER: 36,207  
CC REFERENCE/DOCKET NUMBER: IMI-018CN  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (617)227-7400  
CC TELEFAX: (617)227-5941  
CC INFORMATION FOR SEQ ID NO: 74:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 398 amino acids  
CC TYPE: amino acid  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: protein

SEQUENCE 498 AA: 43664 MW: 779006 CN:

Query Match 51.9% Score 55: DB 1: Length: 398;  
Best Local Similarity 66.7% Pred. No. 3,43e+01;  
Matches 8: Conservative 1: Mismatches 3: Indels 0: Gaps 0:

DB 257 KGLATVAFNMF 268  
1:1 |||||  
QY 1 KSMKVTVAENQF 12

RESULT 9  
ID US-08-175-069A-74 STANDARD: PR: 398 AA.  
XX  
AC xxxxxx  
XX

Sequence 74. Application US/08175069A

Sequence 74, Application US/08175069A  
Patent No. 5776761

GENERAL INFORMATION:

APPLICANT: Rogers, Bruce

APPLICANT: Klapper, David G.

APPLICANT: Rainar, Thorunn

APPLICANT: Kuo, Mei-chang

TITLE OF INVENTION: Allergenic Proteins From Ragweed and Uses

NUMBER OF SEQUENCES: 93

CORRESPONDENCE ADDRESS:

ADDRESSEE: LAHIVE & COCKFIELD, LLP

STREET: 60 State Street

CITY: Boston

STATE: Massachusetts

COUNTRY: USA

ZIP: 02109-1875

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/175,069A

FILING DATE: December 29, 1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/529,951

FILING DATE: May 24, 1990

APPLICATION NUMBER: US 07/525,365

FILING DATE: March 17, 1989

ATTORNEY/AGENT INFORMATION:

NAME: Amy E. Mandragoras

REGISTRATION NUMBER: 36,207

REFERENCE/DOCKET NUMBER: IM-018DV

TELEPHONE: (617)227-7400

TELEFAX: (617)227-5941

INFORMATION FOR SEQ ID NO: 74:

SEQUENCE CHARACTERISTICS:

LENGTH: 398 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE 398 AA: 43664 MW: 779006 CN:

Query Match 51.9% Score 55: DB 1: Length: 398;  
Best Local Similarity 66.7% Pred. No. 3,43e+01;  
Matches 8: Conservative 1: Mismatches 3: Indels 0: Gaps 0:

DB 257 KGLATVAFNMF 268  
1:1 |||||  
QY 1 KSMKVTVAENQF 12

RESULT 10  
ID US-08-290-448A-72 STANDARD: PR: 387 AA.  
XX  
AC xxxxxx  
XX  
DT

Sequence 72. Application US/08290448A

Sequence 72, Application US/08290448A  
Patent No. 5698204

GENERAL INFORMATION:

APPLICANT: Rogers, Bruce

APPLICANT: Klapper, David G.

APPLICANT: Rainar, Thorunn

APPLICANT: Kuo, Mei-chang

TITLE OF INVENTION: Allergenic Proteins From Ragweed and Uses

NUMBER OF SEQUENCES: 93

CORRESPONDENCE ADDRESS:

ADDRESSEE: LAHIVE & COCKFIELD

STREET: 60 State Street, suite 510

CITY: Boston

STATE: Massachusetts

COUNTRY: USA

ZIP: 02109-1875

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/290,448A

FILING DATE: August 15, 1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/529,951

FILING DATE: May 29, 1990

APPLICATION NUMBER: US 07/325,365

FILING DATE: March 17, 1989

ATTORNEY/AGENT INFORMATION:

NAME: Amy E. Mandragoras

REGISTRATION NUMBER: 36,207

REFERENCE/DOCKET NUMBER: IM-018CN

TELEPHONE: (617)227-7400

TELEFAX: (617)227-5941

INFORMATION FOR SEQ ID NO: 72:

SEQUENCE CHARACTERISTICS:

LENGTH: 387 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE 387 AA: 41660 MW: 736967 CN:

Query Match 50.9% Score 54: DB 1: Length 387;  
Best Local Similarity 53.3% Pred. No. 4,37e+01;  
Matches 8: Conservative 2: Mismatches 5: Indels 0: Gaps 0:

DB 246 RGMLATVAFNIFTN 260  
1:1 |||||  
QY 1 KSMKVTVAENQFGPN 15

RESULT 11  
ID US-08-290-448A-72 STANDARD: PR: 387 AA.  
XX  
AC xxxxxx  
XX  
DT

Sequence 72. Application US/08290448A

Sequence 72, Application US/08290448A  
Patent No. 5676954

```

GENERAL INFORMATION:
APPLICANT: Rogers, Bruce
APPLICANT: Klapper, David G.
APPLICANT: Rafnar, Thorunn
APPLICANT: Kuo, Mei-chang
TITLE OF INVENTION: Allergenic Proteins From Raqweed and Uses
NUMBER OF SEQUENCES: 93
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD
STREET: 60 State Street, Suite 510
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109-1875
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/290.443A
FILING DATE: August 15, 1994
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 07/529,951
FILING DATE: May 29, 1990
APPLICATION NUMBER: US 07/325,365
FILING DATE: March 17, 1989
ATTORNEY/AGENT INFORMATION:
NAME: Amy E. Mandragoras
REGISTRATION NUMBER: 36,207
REFERENCE/DOCKET NUMBER: IM1-018CN
TELEPHONE: (617)227-7400
TELEFAX: (617)227-5941
INFORMATION FOR SEQ ID NO: 72:
SEQUENCE CHARACTERISTICS:
LENGTH: 387 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE 387 AA: 41660 MW: 736967 CN:

Query Match 50.98: Score 54: DB: Length 187:
Best Local Similarity 53.38: Pred. No. 4,37e+01:
Matches 8: Conservative 2: Mismatches 5: Gaps 0:

246 KGMIAIVAFNFIIDN 260
1 KSMKVIVAFNFCFN 15

RESULT 12
US-08-175-069A-72 STANDARD: PRI: 387 AA.
xxxxxx
Sequence 72, Application US/08175069A
Patent No. 5775761
GENERAL INFORMATION:
APPLICANT: Rogers, Bruce
APPLICANT: Klapper, David G.
APPLICANT: Rafnar, Thorunn
APPLICANT: Kuo, Mei-chang
TITLE OF INVENTION: Allergenic Proteins From Raqweed and Uses
NUMBER OF SEQUENCES: 93
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD, LLP
STREET: 60 State Street
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109-1875
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/175,069A
FILING DATE: December 29, 1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/529,951
FILING DATE: May 29, 1990
APPLICATION NUMBER: US 07/325,365
FILING DATE: March 17, 1989
ATTORNEY/AGENT INFORMATION:
NAME: Amy E. Mandragoras
REGISTRATION NUMBER: 36,207
REFERENCE/DOCKET NUMBER: IM1-018CN
TELEPHONE: (617)227-7400
TELEFAX: (617)227-5941
INFORMATION FOR SEQ ID NO: 72:
SEQUENCE CHARACTERISTICS:
LENGTH: 387 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE 387 AA: 41660 MW: 736967 CN:

```

```

GENERAL INFORMATION:
APPLICANT: Rogers, Bruce
APPLICANT: Klapper, David G.
APPLICANT: Rafnar, Thorunn
APPLICANT: Kuo, Mei-chang
TITLE OF INVENTION: Allergenic Proteins From Raqweed and Uses
NUMBER OF SEQUENCES: 93
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD, LLP
STREET: 60 State Street
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109-1875
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/175,069A
FILING DATE: December 29, 1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/529,951
FILING DATE: May 29, 1990
APPLICATION NUMBER: US 07/325,365
FILING DATE: March 17, 1989
ATTORNEY/AGENT INFORMATION:
NAME: Amy E. Mandragoras
REGISTRATION NUMBER: 36,207
REFERENCE/DOCKET NUMBER: IM1-018DV
TELEPHONE: (617)227-7400
TELEFAX: (617)227-5941
INFORMATION FOR SEQ ID NO: 72:
SEQUENCE CHARACTERISTICS:
LENGTH: 387 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE 387 AA: 41660 MW: 736967 CN:

Query Match 50.98: Score 54: DB: Length 187:
Best Local Similarity 53.38: Pred. No. 4,37e+01:
Matches 8: Conservative 2: Mismatches 5: Gaps 0:

246 KGMIAIVAFNFIIDN 260
1 KSMKVIVAFNFCFN 15

RESULT 13
US-C8-175-069A-78 STANDARD: PRI: 383 AA.
xxxxxx
Sequence 78, Application US/08175069A
Patent No. 5775761
GENERAL INFORMATION:
APPLICANT: Rogers, Bruce
APPLICANT: Klapper, David G.
APPLICANT: Rafnar, Thorunn
APPLICANT: Kuo, Mei-chang
TITLE OF INVENTION: Allergenic Proteins From Raqweed and Uses
NUMBER OF SEQUENCES: 93
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD, LLP
STREET: 60 State Street
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109-1875
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/175,069A
FILING DATE: December 29, 1993

```

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/529,951  
FILING DATE: May 29, 1990  
APPLICATION NUMBER: US 07/325,365  
FILING DATE: March 17, 1989  
ATTORNEY/AGENT INFORMATION:  
NAME: Amy E. Mandragoras  
REGISTRATION NUMBER: 36,207  
REFERENCE/DOCKET NUMBER: IMI-018CN  
TELEPHONE: (617)227-7400  
TELEFAX: (617)227-7400  
INFORMATION FOR SEQ ID NO: 78:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 383 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE 383 AA: 41852 MW: 738971 CN:

Query Match 50.0% Score 53: DB 1: Length 383:  
Best Local Similarity 53.3% Pred. No. 5.57e+01:  
Matches 8: Conservative 2: Mismatches 5: Indels 0: Gaps 0:

Dz 243 RGMILCTVAFNKFIDN 257  
::: ||||| I  
QY 1 KSMKVIVAFNQFSPN 15

RESULT 14  
ID US-08-290-448A-78 STANDARD: PRT: 383 AA.

AC xxxxxx

Sequence 78, Application US/08290448A

Sequence 78, Application US/08290448A  
Patent No. 5676954

GENERAL INFORMATION:

APPLICANT: Rogers, Bruce  
APPLICANT: Klapper, David G.  
APPLICANT: Rainar, Thorunn  
APPLICANT: Kuo, Mei-Chang  
TITLE OF INVENTION: Antigenic Proteins From Reawood and Uses  
NUMBER OF SEQUENCES: 93  
CORRESPONDENCE ADDRESS:

ADDRESSEE: LAHIVE & COOKFIELD  
STREET: 60 State Street, Suite 510  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02109-1875

COMPUTER READABLE FORM:

COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/290.448A  
FILING DATE: August 15, 1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/529,951  
FILING DATE: May 29, 1990  
APPLICATION NUMBER: US 07/325,365  
FILING DATE: March 17, 1989

ATTORNEY/AGENT INFORMATION:

NAME: Amy E. Mandragoras  
REGISTRATION NUMBER: 36,207  
REFERENCE/DOCKET NUMBER: IMI-018CN  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617)227-7400

TELEFAX: (617)227-5941  
INFORMATION FOR SEQ ID NO: 78:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 383 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE 383 AA: 41852 MW: 738971 CN:

Query Match 50.0% Score 53: DB 1: Length 383:  
Best Local Similarity 53.3% Pred. No. 5.57e+01:  
Matches 8: Conservative 2: Mismatches 5: Indels 0: Gaps 0:

Dz 243 RGMILCTVAFNKFIDN 257  
::: ||||| I  
QY 1 KSMKVIVAFNQFSPN 15

RESULT 15  
ID US-08-290-448A-78 STANDARD: PRT: 383 AA.

AC xxxxxx

Sequence 78, Application US/08290448A

Sequence 78, Application US/08290448A  
Patent No. 5698204

GENERAL INFORMATION:

APPLICANT: Rogers, Bruce  
APPLICANT: Klapper, David G.  
APPLICANT: Rainar, Thorunn  
APPLICANT: Kuo, Mei-Chang  
TITLE OF INVENTION: Antigenic Proteins From Reawood and Uses  
NUMBER OF SEQUENCES: 93  
CORRESPONDENCE ADDRESS:

ADDRESSEE: LAHIVE & COOKFIELD  
STREET: 60 State Street, Suite 510  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02109-1875

COMPUTER READABLE FORM:

COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/290.448A  
FILING DATE: August 15, 1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/529,951  
FILING DATE: May 29, 1990  
APPLICATION NUMBER: US 07/325,365  
FILING DATE: March 17, 1989

ATTORNEY/AGENT INFORMATION:

NAME: Amy E. Mandragoras  
REGISTRATION NUMBER: 36,207  
REFERENCE/DOCKET NUMBER: IMI-018CN  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617)227-7400

TELEFAX: (617)227-7400

INFORMATION FOR SEQ ID NO: 78:

SEQUENCE CHARACTERISTICS:

LENGTH: 383 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE 383 AA: 41852 MW: 738971 CN:

Query Match 50.0% Score 53: DB 1: Length 383:  
Best Local Similarity 53.3% Pred. No. 5.57e+01:



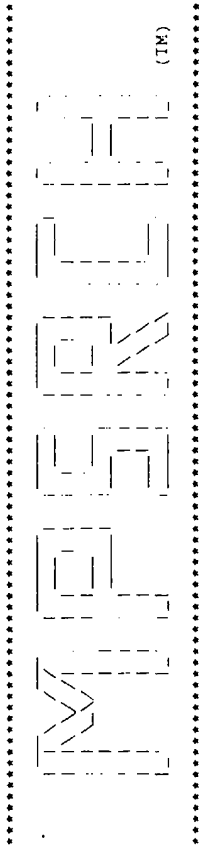
Matches 8: Conservative 2: Mismatches 5: Indels 0: Gaps 0:

PI 243 RSMICTVAENKFTDN 257

III III I I

US 1 KSMKVTVAENKFTDN 15

Search completed: Tue Jun 20 13:46:01 2000  
Job time : 5 secs.



|    |     |       |        |                                   |          |
|----|-----|-------|--------|-----------------------------------|----------|
| 21 | 106 | 100.0 | 28 10  | US-08-467- Sequence 81, Applicat  | 4.81e-05 |
| 22 | 106 | 100.0 | 28 8   | US-08-350- Sequence 219, Applicat | 4.81e-05 |
| 23 | 106 | 100.0 | 28 10  | US-08-468- Sequence 85, Applicat  | 4.81e-05 |
| 24 | 106 | 100.0 | 28 10  | US-08-467- Sequence 82, Applicat  | 4.81e-05 |
| 25 | 106 | 100.0 | 28 10  | US-08-467- Sequence 85, Applicat  | 4.81e-05 |
| 26 | 106 | 100.0 | 28 7   | US-08-228- Sequence 82, Applicat  | 4.81e-05 |
| 27 | 106 | 100.0 | 28 10  | US-08-468- Sequence 81, Applicat  | 4.81e-05 |
| 28 | 106 | 100.0 | 28 10  | US-08-467- Sequence 219, Applicat | 4.81e-05 |
| 29 | 106 | 100.0 | 28 10  | US-08-467- Sequence 82, Applicat  | 4.81e-05 |
| 30 | 106 | 100.0 | 29 10  | US-08-467- Sequence 222, Applicat | 4.81e-05 |
| 31 | 106 | 100.0 | 30 10  | US-08-467- Sequence 79, Applicat  | 4.81e-05 |
| 32 | 106 | 100.0 | 30 7   | US-08-228- Sequence 80, Applicat  | 4.81e-05 |
| 33 | 106 | 100.0 | 30 10  | US-08-467- Sequence 80, Applicat  | 4.81e-05 |
| 34 | 106 | 100.0 | 30 10  | US-08-468- Sequence 80, Applicat  | 4.81e-05 |
| 35 | 106 | 100.0 | 30 10  | US-08-467- Sequence 83, Applicat  | 4.81e-05 |
| 36 | 106 | 100.0 | 30 10  | US-08-467- Sequence 78, Applicat  | 4.81e-05 |
| 37 | 106 | 100.0 | 36 10  | US-08-467- Sequence 78, Applicat  | 4.81e-05 |
| 38 | 106 | 100.0 | 36 10  | US-08-467- Sequence 78, Applicat  | 4.81e-05 |
| 39 | 106 | 100.0 | 90 10  | US-08-468- Sequence 54, Applicat  | 4.81e-05 |
| 40 | 106 | 100.0 | 90 4   | US-07-938- Sequence 54, Applicat  | 4.81e-05 |
| 41 | 106 | 100.0 | 90 10  | US-08-467- Sequence 54, Applicat  | 4.81e-05 |
| 42 | 106 | 100.0 | 367 10 | US-08-467- Sequence 95, Applicat  | 4.81e-05 |
| 43 | 106 | 100.0 | 367 10 | US-08-467- Sequence 95, Applicat  | 4.81e-05 |
| 44 | 106 | 100.0 | 374 8  | US-08-350- Sequence 2, Applicatio | 4.81e-05 |
| 45 | 106 | 100.0 | 374 4  | US-07-938- Sequence 2, Applicatio | 4.81e-05 |

ALIGNMENTS

RESULT 1  
ID US-08-467-023-227 STANDARD: PRT: 19 AA.  
XX  
AC xxxxxx  
XX  
DT  
XX  
XX  
DE Sequence 227, Application US/08467023  
XX  
CC Sequence 227, Application US/08467023  
CC GENERAL INFORMATION:  
CC APPLICANT: Griffith, Edwin J.  
CC APPLICANT: Pollock, Joanne.  
CC APPLICANT: Reed, Julian P.  
CC APPLICANT: Garman, Richard P.  
CC APPLICANT: Koo, Melcham.  
CC APPLICANT: Yeh, Siguang H.  
CC APPLICANT: Brauer, Andrew.  
CC APPLICANT: Exley, Mark A.  
CC APPLICANT: Powels, Steven P.  
CC TITLE OF INVENTION: Allergenic Proteins And Peptides From  
CC TITLE OF INVENTION: Japanese Cedar Pollen  
CC NUMBER OF SEQUENCES: 261  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Immunologic Pharmaceutical Corporation, Inc.  
CC STREET: 610 Lincoln St  
CC CITY: Waltham  
CC STATE: MA  
CC COUNTRY: USA  
CC ZIP: 02154  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: Patent Release #1.0, Version #1.25  
CC CURRENT APPLICATION NUMBER: US/08/467,023  
CC FILING DATE: June 6, 1995  
CC CLASSIFICATION: 424  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: 08/350,225  
CC FILING DATE: December 6, 1994  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Jane E. Remilla-J

REGISTRATION NUMBER: 38,872  
 REFERENCE/DOCKET NUMBER: 025.6 USD2 (IM-028CPD2)  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (617) 227-7400  
 TELEFAX: (617) 227-5441  
 INFORMATION FOR SEQ ID NO: 227:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 19 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: peptide  
 FRAGMENT TYPE: internal  
 SEQUENCE: 19 AA: 2156 MW: 1801 CN:  
 Query Match: 100.0% Score 100.00 E-010 Length 19:  
 Best Local Similarity 100.0% Prod. No. 4.8.e-05:  
 Matches 15: Conservative 0: Mismatches 0: Indels 0: Gaps 0:  
 Db 3 KSMKVIVAFNFGPN 17  
 QY 1 KSMKVIVAFNFGPN 15  
 RESULT 2  
 ID US-09-350-225-227 STANDARD: PRI: 19 AA:  
 XX xxxxxx  
 Sequence 227, Application US/08350225  
 Sequence 227, Application US/08350225  
 GENERAL INFORMATION:  
 APPLICANT: Griffith, Irwin J.  
 APPLICANT: Pollock, Joanne  
 APPLICANT: Bond, Julian F.  
 APPLICANT: Gattac, Richard D.  
 APPLICANT: Kuo, Mei-Chang  
 APPLICANT: Yeung, Siu-mei H.  
 APPLICANT: Brauer, Andrew  
 APPLICANT: Exley, Mark A.  
 APPLICANT: Powers, Steven P.  
 TITLE OF INVENTION: Allergenic Proteins And Peptides From  
 TITLE OF INVENTION: Japanese Cedar Pollen  
 NUMBER OF SEQUENCES: 2  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Immunogen Pharmaceuticals Corporation, Inc.  
 STREET: 610 Lincoln St.  
 CITY: Waltham  
 STATE: MA  
 COUNTRY: USA  
 ZIP: 02154  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: F.COPY disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/350,225  
 FILING DATE: December 6, 1994  
 CLASSIFICATION: 424  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 08/225,248  
 FILING DATE: April 8, 1994  
 APPLICATION NUMBER: 07/948,990  
 FILING DATE: September 1, 1992  
 APPLICATION NUMBER: PC/US93/00139  
 FILING DATE: January 15, 1993  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Carlene A. Varscone  
 REGISTRATION NUMBER: 35,729  
 REFERENCE/DOCKET NUMBER: 025.6 US (IM-028CP2)

TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (617) 466-6000  
 TELEFAX: (617) 466-6240  
 INFORMATION FOR SEQ ID NO: 227:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 19 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: peptide  
 FRAGMENT TYPE: internal  
 SEQUENCE: 19 AA: 2156 MW: 1801 CN:  
 Query Match: 100.0% Score 100.00 E-010 Length 19:  
 Best Local Similarity 100.0% Prod. No. 4.8.e-05:  
 Matches 15: Conservative 0: Mismatches 0: Indels 0: Gaps 0:  
 Db 3 KSMKVIVAFNFGPN 17  
 QY 1 KSMKVIVAFNFGPN 15  
 RESULT 3  
 ID US-09-467-023-230 STANDARD: PRI: 2: AA:  
 XX xxxxxx  
 Sequence 210, Application US/08467023  
 Sequence 230, Application US/08467023  
 GENERAL INFORMATION:  
 APPLICANT: Griffith, Irwin J.  
 APPLICANT: Pollock, Joanne  
 APPLICANT: Bond, Julian F.  
 APPLICANT: Gattac, Richard D.  
 APPLICANT: Kuo, Mei-Chang  
 APPLICANT: Yeung, Siu-mei H.  
 APPLICANT: Brauer, Andrew  
 APPLICANT: Exley, Mark A.  
 APPLICANT: Powers, Steven P.  
 TITLE OF INVENTION: Allergenic Proteins And Peptides From  
 TITLE OF INVENTION: Japanese Cedar Pollen  
 NUMBER OF SEQUENCES: 241  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Immunogen Pharmaceuticals Corporation, Inc.  
 STREET: 610 Lincoln St.  
 CITY: Waltham  
 STATE: MA  
 COUNTRY: USA  
 ZIP: 02154  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/467,023  
 FILING DATE: June 6, 1995  
 CLASSIFICATION: 424  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 08/350,225  
 FILING DATE: December 6, 1994  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Jane E. Remillard  
 REGISTRATION NUMBER: 38,872  
 REFERENCE/DOCKET NUMBER: 025.6 USD2 (IM-028CP22)  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (617) 227-7400  
 TELEFAX: (617) 227-5441  
 INFORMATION FOR SEQ ID NO: 230:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 20 amino acids

CC TYPE: amino acid  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: peptide  
CC FRAGMENT TYPE: internal  
SQ SEQUENCE 20 AA: 2285 MW: 2007 CN:

Query Match 100.0% Score 106; DB 10; Length 20;  
Best Local Similarity 100.0% Pred. No. 4.81e-05;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 4 KSMKVTVAFNQFGPN 18  
QY 1 KSMKVTVAFNQFGPN 15

RESULT 4  
ID US-07-938-953A-47 STANDARD: PRT: 20 AA.

XX XXXXXX

Sequence 47, Application US/07938990A

Sequence 47, Application US/07938990A

GENERAL INFORMATION:

APPLICANT: Griffith, Irwin J.

APPLICANT: Pollock, Joanne

APPLICANT: Bond, Julian

TITLE OF INVENTION: Allergenic Proteins And Peptides From

TITLE OF INVENTION: Japanese Cedar Pollen

NUMBER OF SEQUENCES: 70

CORRESPONDENCE ADDRESS:

ADDRESSES: Lahive & Cockfield

STREET: Sixty State Street

CITY: Boston

STATE: MA

COUNTRY: USA

ZIP: 02109

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/948,990A

FILING DATE: 1992-09-01

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 07/730,452

FILING DATE: July 15, 1991

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 07/739,134

FILING DATE: July 10, 1991

ATTORNEY/AGENT INFORMATION:

NAME: Amy E. Mandragoras

REGISTRATION NUMBER: 36,207

REFERENCE/DOCKET NUMBER: IPC-C25CC (IMI-028)

TELEPHONE: (617) 227-7400

TELEFAX: (617) 227-5941

INFORMATION FOR SEQ ID NO: 47:

SEQUENCE CHARACTERISTICS:

LENGTH: 20 amino acids

TYPE: AMINO ACID

TOPOLOGY: linear

MOLECULE TYPE: peptide

FRAGMENT TYPE: internal

SEQUENCE 20 AA: 2244 MW: 1852 CN:

Query Match 100.0% Score 106; DB 4; Length 20;  
Best Local Similarity 100.0% Pred. No. 4.81e-05;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 1 KSMKVTVAFNQFGPN 15  
QY 1 KSMKVTVAFNQFGPN 15

RESULT 5  
ID US-08-350-225-230 STANDARD: PRT: 20 AA.

XX XXXXXX

Sequence 230, Application US/08350225

Sequence 230, Application US/08350225

GENERAL INFORMATION:

APPLICANT: Griffith, Irwin J.

APPLICANT: Pollock, Joanne

APPLICANT: Bond, Julian

APPLICANT: Garman, Richard D.

APPLICANT: Kuo, Mei-Chang

APPLICANT: Yeung, Siu-mei H.

APPLICANT: Brauer, Andrew

APPLICANT: Exley, Mark A.

APPLICANT: Powers, Steven P.

TITLE OF INVENTION: Allergenic Proteins And Peptides From

TITLE OF INVENTION: Japanese Cedar Pollen

NUMBER OF SEQUENCES: 261

CORRESPONDENCE ADDRESS:

ADDRESSEE: Immunologic Pharmaceutical Corporation, Inc.

STREET: 510 Lincoln St

CITY: Waltham

STATE: MA

COUNTRY: USA

ZIP: 02154

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/350,225

FILING DATE: December 1, 1994

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/221,248

FILING DATE: April 8, 1994

APPLICATION NUMBER: 07/945,990

FILING DATE: September 1, 1992

APPLICATION NUMBER: PCI/US93/00139

FILING DATE: January 15, 1993

ATTORNEY/AGENT INFORMATION:

NAME: Darlene A. Vanstone

REGISTRATION NUMBER: 35,729

REFERENCE/DOCKET NUMBER: 025.6 US (IMI-028CP2)

TELEPHONE: (617) 466-6700

TELEFAX: (617) 466-604

INFORMATION FOR SEQ ID NO: 230:

SEQUENCE CHARACTERISTICS:

LENGTH: 20 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: peptide

FRAGMENT TYPE: internal

SEQUENCE 20 AA: 2285 MW: 2007 CN:

Query Match 100.0% Score 106; DB 8; Length 20;  
Best Local Similarity 100.0% Pred. No. 4.81e-05;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 4 KSMKVTVAFNQFGPN 18

1 KSMKVTVAFNFGPN 15

US-08-467-697-47 STANDARD: PRI: 20 AA.

xxxxxx

Sequence 47, Application US/08457697

Sequence 47, Application US/08457697

GENERAL INFORMATION:  
 APPLICANT: Griffith, Irwin J.  
 APPLICANT: Pollock, Joanne  
 APPLICANT: Bond, Julia F.  
 APPLICANT: Garman, Richard D.  
 APPLICANT: Kuo, Mei-Chang  
 APPLICANT: Yeung, Siu-mei H.  
 APPLICANT: Brauer, Andrew  
 APPLICANT: Exley, Mark A.  
 APPLICANT: Powers, Steven P.  
 TITLE OF INVENTION: Allergenic Proteins And Peptides From  
 TITLE OF INVENTION: Japanese Cedar Pollen  
 NUMBER OF SEQUENCES: 261  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Immunologic Pharmaceutical Corporation, Inc.  
 STREET: 610 Lincoln St  
 CITY: Waltham  
 STATE: MA  
 COUNTRY: USA  
 ZIP: 02154

COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent In Release #1.0, Version #1.25  
 CURRENT APPLICATION NUMBER: US/08/467,697  
 FILING DATE: June 6, 1995  
 CLASSIFICATION: 424  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 08/350,225  
 FILING DATE: December 6, 1994  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Jane E. Remillard  
 REGISTRATION NUMBER: 98,972  
 REFERENCE/DOCKET NUMBER: 025.6 US/4 (IMI-0282704)  
 TELEPHONE: (617) 227-7400  
 TELEFAX: (617) 227-5941  
 INFORMATION FOR SEQ ID NO: 47:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 20 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: peptide  
 FRAGMENT TYPE: internal

SEQUENCE 20 AA: 2244 MW: 1852 CN:

Query Match 100.0% Score 106; DB 10; Length 20;  
 Best Local Similarity 100.0%; Pred. No. 4,81e-05;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 KSMKVTVAFNFGPN 15

1 KSMKVTVAFNFGPN 15

1 KSMKVTVAFNFGPN 15

1 KSMKVTVAFNFGPN 15

1 KSMKVTVAFNFGPN 15

1 KSMKVTVAFNFGPN 15

1 KSMKVTVAFNFGPN 15

1 KSMKVTVAFNFGPN 15

1 KSMKVTVAFNFGPN 15

1 KSMKVTVAFNFGPN 15

1 KSMKVTVAFNFGPN 15

1 KSMKVTVAFNFGPN 15

1 KSMKVTVAFNFGPN 15

xxxxxx

Sequence 231, Application US/08467023

Sequence 231, Application US/08467023

GENERAL INFORMATION:  
 APPLICANT: Griffith, Irwin J.  
 APPLICANT: Pollock, Joanne  
 APPLICANT: Bond, Julia F.  
 APPLICANT: Garman, Richard D.  
 APPLICANT: Kuo, Mei-Chang  
 APPLICANT: Yeung, Siu-mei H.  
 APPLICANT: Brauer, Andrew  
 APPLICANT: Exley, Mark A.  
 APPLICANT: Powers, Steven P.  
 TITLE OF INVENTION: Allergenic Proteins And Peptides From  
 TITLE OF INVENTION: Japanese Cedar Pollen  
 NUMBER OF SEQUENCES: 261  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Immunologic Pharmaceutical Corporation, Inc.  
 STREET: 610 Lincoln St  
 CITY: Waltham  
 STATE: MA  
 COUNTRY: USA  
 ZIP: 02154

COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent In Release #1.0, Version #1.25  
 CURRENT APPLICATION NUMBER: US/08/467,697  
 FILING DATE: June 6, 1995  
 CLASSIFICATION: 424  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 08/350,225  
 FILING DATE: December 6, 1994  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Jane E. Remillard  
 REGISTRATION NUMBER: 98,972  
 REFERENCE/DOCKET NUMBER: 025.6 US/4 (IMI-0282704)  
 TELEPHONE: (617) 227-7400  
 TELEFAX: (617) 227-5941  
 INFORMATION FOR SEQ ID NO: 47:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 21 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: peptide  
 FRAGMENT TYPE: internal

SEQUENCE 21 AA: 2415 MW: 2154 CN:

Query Match 100.0% Score 106; DB 10; Length 21;  
 Best Local Similarity 100.0%; Pred. No. 4,81e-05;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

4 KSMKVTVAFNFGPN 18

1 KSMKVTVAFNFGPN 15

1 KSMKVTVAFNFGPN 15

1 KSMKVTVAFNFGPN 15

1 KSMKVTVAFNFGPN 15

1 KSMKVTVAFNFGPN 15

1 KSMKVTVAFNFGPN 15

1 KSMKVTVAFNFGPN 15

1 KSMKVTVAFNFGPN 15

1 KSMKVTVAFNFGPN 15

1 KSMKVTVAFNFGPN 15

1 KSMKVTVAFNFGPN 15

1 KSMKVTVAFNFGPN 15

1 KSMKVTVAFNFGPN 15

1 KSMKVTVAFNFGPN 15

1 KSMKVTVAFNFGPN 15

1 KSMKVTVAFNFGPN 15

1 KSMKVTVAFNFGPN 15

```

XX Sequence 231, Application US/08457006
CC GENERAL INFORMATION:
CC APPLICANT: Griffith, Irwin J.:
CC APPLICANT: Pollock, Joanne:
CC APPLICANT: Bond, Julian F.:
CC APPLICANT: Garman, Richard D:
CC APPLICANT: Kuo, Mei-Cha:
CC APPLICANT: Yeung, Siu-mei H.:
CC APPLICANT: Brauer, Andrew:
CC APPLICANT: Exley, Mark A.:
CC APPLICANT: Powers, Steven P:
CC APPLICANT: Griffith, Irwin J.:
CC APPLICANT: Pollock, Joanne:
CC APPLICANT: Bond, Julian F.:
CC TITLE OF INVENTION: Allergenic Proteins And Peptides From
CC TITLE OF INVENTION: Japanese Cedar Pollen
CC NUMBER OF SEQUENCES: 261
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Immunologic Pharmaceutical Corporation, Inc.
CC STREET: 610 Lincoln St
CC CITY: Waltham
CC STATE: MA
CC COUNTRY: USA
CC ZIP: 02154
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patent Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08457.006
CC FILING DATE: June 6, 1995
CC CLASSIFICATION: 424
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/350,225
CC FILING DATE: December 6, 1994
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Zace E. Remillard
CC REGISTRATION NUMBER: 38,872
CC REFERENCE/DOCKET NUMBER: 025.6 US 5 (IM: 028CPD5)
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (617) 227-7400
CC TELEFAX: (617) 227-5941
CC INFORMATION FOR SEQ ID NO: 231:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 21 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC FRAGMENT TYPE: internal
CC SEQUENCE 21 AA: 2415 KW: 2154 CN:
SQ

```

Query Match: 100.0%; Score 106; DB 10; Length 21;  
 Best Local Similarity 100.0%; Pred. No. 4.81e-05;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

D1 4 KSMKVTVAFNQFGPN 18
Q2 1 KSMKVTVAFNQFGPN 15

```

```

RESULT 9
ID US-08-468-940-233 STANDARD: PRI: 23 AA.
XX
AC xxxxxx
XX
DT
XX
DE Sequence 233, Application US/08468940
XX
CC Sequence 233, Application US/08468940
CC GENERAL INFORMATION:
CC APPLICANT: Griffith, Irwin J.:
CC APPLICANT: Pollock, Joanne:
CC APPLICANT: Bond, Julian F.:

```

```

CC APPLICANT: Garman, Rich : D:
CC APPLICANT: Kuo, Mei-Cha:
CC APPLICANT: Yeung, Siu-mei H.:
CC APPLICANT: Brauer, Andrew:
CC APPLICANT: Exley, Mark A.:
CC APPLICANT: Powers, Steven P:
CC TITLE OF INVENTION: Allergenic Proteins And Peptides From
CC TITLE OF INVENTION: Japanese Cedar Pollen
CC NUMBER OF SEQUENCES: 261
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Immunologic Pharmaceutical Corporation, Inc.
CC STREET: 610 Lincoln St
CC CITY: Waltham
CC STATE: MA
CC COUNTRY: USA
CC ZIP: 02154
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patent Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08458.940
CC FILING DATE:
CC CLASSIFICATION: 424
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/350,225
CC FILING DATE: December 6, 1994
CC APPLICATION NUMBER: 08/226,248
CC FILING DATE: April 8, 1994
CC APPLICATION NUMBER: 07/936,990
CC FILING DATE: September 1, 1992
CC APPLICATION NUMBER: PCI/US93/00134
CC FILING DATE: January 15, 1993
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Darlene A. Vanstone
CC REGISTRATION NUMBER: 35,729
CC REFERENCE/DOCKET NUMBER: 025.6 US (IM: 028CP2);
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (617) 466-5000
CC TELEFAX: (617) 466-6040
CC INFORMATION FOR SEQ ID NO: 233:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 23 amino acid
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC FRAGMENT TYPE: internal
CC SEQUENCE 23 AA: 2558 MW: 2164 CN:
SQ

```

Query Match: 100.0%; Score 106; DB 10; Length 23;  
 Best Local Similarity 100.0%; Pred. No. 4.81e-05;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Db 4 KSMKVTVAFNQFGPN 18
Qy 1 KSMKVTVAFNQFGPN 15

```

```

RESULT 10
ID US-08-467-023-224 STANDARD: PRI: 24 AA.
XX
AC xxxxxx
XX
DT
XX
DE Sequence 224, Application US/08467023
XX
CC Sequence 224, Application US/08467023
CC GENERAL INFORMATION:
CC APPLICANT: Griffith, Irwin J.:
CC APPLICANT: Pollock, Joanne:
CC APPLICANT: Bond, Julian F.:

```

CC APPLICANT: Garman, Richard D.  
 CC APPLICANT: Kuo, Mei-Chang;  
 CC APPLICANT: Yeung, Siu-mei H.;  
 CC APPLICANT: Brauer, Andrew;  
 CC APPLICANT: Exley, Mark A.;  
 CC APPLICANT: Powers, Steven P.  
 CC TITLE OF INVENTION: Allergenic Proteins And Peptides From  
 CC TITLE OF INVENTION: Japanese Cedar Pollen  
 CC NUMBER OF SEQUENCES: 26;  
 CC CORRESPONDENCE ADDRESS:  
 CC ADDRESSEE: Immunologic Pharmaceutical Corporation, Inc.  
 CC STREET: 610 Lincoln St  
 CC CITY: Waltham  
 CC STATE: MA  
 CC COUNTRY: USA  
 CC ZIP: 02154  
 CC COMPUTER READABLE FORM:  
 CC MEDIUM TYPE: Floppy disk  
 CC COMPUTER: IBM PC compatible  
 CC OPERATING SYSTEM: PC-DOS/MS-DOS  
 CC SOFTWARE: Patent In Release #1.0, Version #1.25  
 CC CURRENT APPLICATION DATA:  
 CC APPLICATION NUMBER: US/08/467,023  
 CC FILING DATE: June 6, 1995  
 CC CLASSIFICATION: 424  
 CC PRIOR APPLICATION DATA:  
 CC APPLICATION NUMBER: 08/350,225  
 CC FILING DATE: December 5, 1994  
 CC ATTORNEY/AGENT INFORMATION:  
 CC NAME: Jace E. Renillard  
 CC REGISTRATION NUMBER: 38,872  
 CC REFERENCE/DOCKET NUMBER: 025.6 US22 (IMI-028CPD2)  
 CC TELECOMMUNICATION INFORMATION:  
 CC TELEPHONE: (617) 227-7400  
 CC TELEFAX: (617) 227-5941  
 CC INFORMATION FOR SEQ ID NO: 224:  
 CC SEQUENCE CHARACTERISTICS:  
 CC LENGTH: 24 amino acids  
 CC TYPE: amino acid  
 CC TOPOLOGY: linear  
 CC MOLECULE TYPE: peptide  
 CC FRAGMENT TYPE: internal  
 CC SEQUENCE 24 AA: 2716 MW: 2744 CN:  
 SQ

Query Match 100.0% Score 105. DB 10 Length 24  
 Best Local Similarity 100.0% Pred. No. 4.81e-05  
 Matches 15 Conservative 83 Mismatches 0 Labels 00 Gaps 00

DL 3 KSKKVTVAFFNGFQPN 17  
 U 1 KSKKVTVAFFNGFQPN 15

RESULT 11  
 ID US-08-468-940-86 STANDARD: PRI: 26 AA.  
 XX  
 AC xxxxxx  
 DT  
 XX  
 DE  
 XX

Sequence 86, Application US/08458940  
 GENERAL INFORMATION:  
 CC APPLICANT: Griffith, Irwin J.;  
 CC APPLICANT: Pollock, Joanne;  
 CC APPLICANT: Bond, Julian F.;  
 CC APPLICANT: Garman, Richard D.;  
 CC APPLICANT: Kuo, Mei-Chang;  
 CC APPLICANT: Yeung, Siu-mei H.;  
 CC APPLICANT: Brauer, Andrew;  
 CC APPLICANT: Exley, Mark A.;  
 CC APPLICANT: Powers, Steven P.

CC TITLE OF INVENTION: Allergenic Proteins And Peptides From  
 CC TITLE OF INVENTION: Japanese Cedar Pollen  
 CC NUMBER OF SEQUENCES: 261  
 CC CORRESPONDENCE ADDRESS:  
 CC ADDRESSEE: Immunologic Pharmaceutical Corporation, Inc.  
 CC STREET: 610 Lincoln St  
 CC CITY: Waltham  
 CC STATE: MA  
 CC COUNTRY: USA  
 CC ZIP: 02154  
 CC COMPUTER READABLE FORM:  
 CC MEDIUM TYPE: Floppy disk  
 CC COMPUTER: IBM PC compatible  
 CC OPERATING SYSTEM: PC-DOS/MS-DOS  
 CC SOFTWARE: Patent In Release #1.0, Version #1.25  
 CC CURRENT APPLICATION DATA:  
 CC APPLICATION NUMBER: US/08/458,940  
 CC FILING DATE:  
 CC CLASSIFICATION: 424  
 CC PRIOR APPLICATION DATA:  
 CC APPLICATION NUMBER: 08/350,225  
 CC FILING DATE: December 5, 1994  
 CC APPLICATION NUMBER: 08/226,248  
 CC FILING DATE: April 8, 1994  
 CC APPLICATION NUMBER: 07/48,990  
 CC FILING DATE: September 1, 1992  
 CC APPLICATION NUMBER: 01 US93/00139  
 CC FILING DATE: January 15, 1993  
 CC ATTORNEY/AGENT INFORMATION:  
 CC NAME: Darlene A. Vanstone  
 CC REGISTRATION NUMBER: 35,729  
 CC REFERENCE/DOCKET NUMBER: 025.6 US (IMI-028CP2)  
 CC TELECOMMUNICATION INFORMATION:  
 CC TELEPHONE: (617) 466-6400  
 CC TELEFAX: (617) 466-6040  
 CC INFORMATION FOR SEQ ID NO: 86:  
 CC SEQUENCE CHARACTERISTICS:  
 CC LENGTH: 26 amino acids  
 CC TYPE: amino acid  
 CC TOPOLOGY: linear  
 CC MOLECULE TYPE: peptide  
 CC FRAGMENT TYPE: internal  
 CC SEQUENCE 26 AA: 2944 MW: 2941 CN:  
 SQ

Query Match 100.0% Score 105. DB 10 Length 26  
 Best Local Similarity 100.0% Pred. No. 1.40e-06  
 Matches 15 Conservative 83 Mismatches 0 Labels 00 Gaps 00

DL 1 KSKKVTVAFFNGFQPN 15  
 U 1 KSKKVTVAFFNGFQPN 15

RESULT 12  
 ID US-08-468-940-83 STANDARD: PRI: 26 AA.  
 XX  
 AC xxxxxx  
 DT  
 XX  
 DE  
 XX

Sequence 83, Application US/08458940  
 GENERAL INFORMATION:  
 CC APPLICANT: Griffith, Irwin J.;  
 CC APPLICANT: Pollock, Joanne;  
 CC APPLICANT: Bond, Julian F.;  
 CC APPLICANT: Garman, Richard D.;  
 CC APPLICANT: Kuo, Mei-Chang;  
 CC APPLICANT: Yeung, Siu-mei H.;  
 CC APPLICANT: Brauer, Andrew;  
 CC APPLICANT: Exley, Mark A.;  
 CC APPLICANT: Powers, Steven P.

CC TITLE OF INVENTION: Allergenic Proteins And Peptides From  
 CC TITLE OF INVENTION: Japanese Cedar Pollen  
 CC NUMBER OF SEQUENCES: 261  
 CC CORRESPONDENCE ADDRESS:  
 CC ADDRESSEE: ImmuLogic Pharmaceutical Corporation, Inc.  
 CC STREET: 610 Lincoln St  
 CC CITY: Waltham  
 CC STATE: MA  
 CC COUNTRY: USA  
 CC ZIP: 02154  
 CC COMPUTER READABLE FORM:  
 CC MEDIUM TYPE: Floppy disk  
 CC COMPUTER: IBM PC compatible  
 CC OPERATING SYSTEM: PC-DOS/MS-DOS  
 CC SOFTWARE: PatentIn Release #1.0, Version #1.25  
 CC CURRENT APPLICATION DATA:  
 CC APPLICATION NUMBER: US/08/468,940  
 CC FILING DATE:  
 CC CLASSIFICATION: 424  
 CC PRIOR APPLICATION DATA:  
 CC APPLICATION NUMBER: 08/350,225  
 CC FILING DATE: December 6, 1994  
 CC APPLICATION NUMBER: 08/226,248  
 CC FILING DATE: April 8, 1994  
 CC APPLICATION NUMBER: 07/338,990  
 CC FILING DATE: September 11, 1992  
 CC APPLICATION NUMBER: PC/US93/00139  
 CC FILING DATE: January 15, 1993  
 CC ATTORNEY/AGENT INFORMATION:  
 CC NAME: Darlene A. Vanstone  
 CC REGISTRATION NUMBER: 35,729  
 CC REFERENCE/DOCKET NUMBER: 025.6 US (IMI-028C22)  
 CC TELECOMMUNICATION INFORMATION:  
 CC TELEPHONE: (617) 466-6000  
 CC TELEFAX: (617) 466-6040  
 CC INFORMATION FOR SEQ ID NO: 83:  
 CC SEQUENCE CHARACTERISTICS:  
 CC LENGTH: 26 amino acids  
 CC TYPE: amino acid  
 CC TOPOLOGY: linear  
 CC MOLECULE TYPE: peptide  
 CC FRAGMENT TYPE: internal  
 CC SEQUENCE 26 AA: 2944 MW: 2965 CN  
 SQ  
 Query Match 100.0% Score 104. LR 1 Length 26  
 Best Local Similarity 100.0% Pred. No. 4,81e-05  
 Matches 15: Conservative 0: Mismatches 0: Indels 0: Gaps 0:  
 DB 1 KSMKVTVAENQFGPN 15  
 QY :|||||:|||||  
 RESULT 13  
 ID US-08-226-248A-93 STANDARD: PRI: 25 AA:  
 XX xxxxxx  
 AC  
 XX  
 DT  
 DE  
 Sequence 83, Application US/08226248A  
 Sequence 83, Application US/08226248A  
 GENERAL INFORMATION:  
 CC APPLICANT: Griffith, Irwin J.:  
 CC APPLICANT: Pollock, Joanne:  
 CC APPLICANT: Bond, Julian F.:  
 CC APPLICANT: Garman, Richard D:  
 CC APPLICANT: Kuo, Mei-Chang:  
 CC APPLICANT: Yeung, Siu-mei H.:  
 CC APPLICANT: Brauer, Andrew:  
 CC APPLICANT: Exley, Mark A.:  
 CC APPLICANT: Powers, Steven P.

CC TITLE OF INVENTION: Allergenic Proteins And Peptides From  
 CC TITLE OF INVENTION: Japanese Cedar Pollen  
 CC NUMBER OF SEQUENCES: 261  
 CC CORRESPONDENCE ADDRESS:  
 CC ADDRESSEE: ImmuLogic Pharmaceutical Corporation, Inc.  
 CC STREET: 610 Lincoln St  
 CC CITY: Waltham  
 CC STATE: MA  
 CC COUNTRY: USA  
 CC ZIP: 02154  
 CC COMPUTER READABLE FORM:  
 CC MEDIUM TYPE: Floppy disk  
 CC COMPUTER: IBM PC compatible  
 CC OPERATING SYSTEM: PC-DOS/MS-DOS  
 CC SOFTWARE: PatentIn Release #1.0, Version #1.25  
 CC CURRENT APPLICATION DATA:  
 CC APPLICATION NUMBER: US/08/226,248A  
 CC FILING DATE: April 8, 1994  
 CC CLASSIFICATION: 424  
 CC PRIOR APPLICATION DATA:  
 CC APPLICATION NUMBER: 07/518,460  
 CC FILING DATE: September 11, 1992  
 CC APPLICATION NUMBER: PC/US93/00139  
 CC FILING DATE: January 15, 1993  
 CC ATTORNEY/AGENT INFORMATION:  
 CC NAME: Darlene A. Vanstone  
 CC REGISTRATION NUMBER: 35,729  
 CC REFERENCE/DOCKET NUMBER: 025.5 US (IMI-028CP)  
 CC TELECOMMUNICATION INFORMATION:  
 CC TELEPHONE: (617) 466-6000  
 CC TELEFAX: (617) 466-6040  
 CC INFORMATION FOR SEQ ID NO: 83:  
 CC SEQUENCE CHARACTERISTICS:  
 CC LENGTH: 26 amino acids  
 CC TYPE: amino acid  
 CC TOPOLOGY: linear  
 CC MOLECULE TYPE: peptide  
 CC FRAGMENT TYPE: internal  
 CC SEQUENCE 26 AA: 2944 MW: 2965 CN  
 SQ  
 Query Match 100.0% Score 104. LR 1 Length 26  
 Best Local Similarity 100.0% Pred. No. 4,81e-05  
 Matches 15: Conservative 0: Mismatches 0: Indels 0: Gaps 0:  
 DB 1 KSMKVTVAENQFGPN 15  
 QY :|||||:|||||  
 RESULT 14  
 ID US-08-467-697-86 STANDARD: PRI: 26 AA:  
 XX xxxxxx  
 AC  
 XX  
 DT  
 DE  
 Sequence 86, Application US/08467697  
 Sequence 86, Application US/08467697  
 GENERAL INFORMATION:  
 CC APPLICANT: Griffith, Irwin J.:  
 CC APPLICANT: Pollock, Joanne:  
 CC APPLICANT: Bond, Julian F.:  
 CC APPLICANT: Garman, Richard D:  
 CC APPLICANT: Kuo, Mei-Chang:  
 CC APPLICANT: Yeung, Siu-mei H.:  
 CC APPLICANT: Brauer, Andrew:  
 CC APPLICANT: Exley, Mark A.:  
 CC APPLICANT: Powers, Steven P.  
 CC TITLE OF INVENTION: Allergenic Proteins And Peptides From  
 CC NUMBER OF SEQUENCES: 261  
 CC CORRESPONDENCE ADDRESS:



CC ADDRESSEE: Immunologic Pharmaceutical Corporation, Inc.  
 CC STREET: 610 Lincoln St  
 CC CITY: Waltham  
 CC STATE: MA  
 CC COUNTRY: USA  
 CC ZIP: 02154  
 CC COMPUTER READABLE FORM:  
 CC MEDIUM TYPE: Floppy disk  
 CC COMPUTER: IBM PC compatible  
 CC OPERATING SYSTEM: PC-DOS/MS-DOS  
 CC SOFTWARE: Patent Release #1.0, Version #1.25  
 CC CURRENT APPLICATION DATA:  
 CC APPLICATION NUMBER: US/08/467,697  
 CC FILING DATE: June 6, 1995  
 CC CLASSIFICATION: 424  
 CC PRIOR APPLICATION DATA:  
 CC APPLICATION NUMBER: 08/350,225  
 CC FILING DATE: December 6, 1994  
 CC ATTORNEY/AGENT INFORMATION:  
 CC NAME: Jane E. Remillard  
 CC REGISTRATION NUMBER: 38,872  
 CC REFERENCE/DOCKET NUMBER: 025.6 USD4 (IMI-028CPD4);  
 CC TELECOMMUNICATION INFORMATION:  
 CC TELEPHONE: (617) 227-7400  
 CC TELEFAX: (617) 227-5941  
 CC INFORMATION FOR SEQ ID NO: 86:  
 CC SEQUENCE CHARACTERISTICS:  
 CC LENGTH: 26 amino acids  
 CC TYPE: amino acid  
 CC TOPOLOGY: linear  
 CC MOLECULE TYPE: peptide  
 CC FRAGMENT TYPE: internal  
 CC SEQUENCE 26 AA: 2944 MW: 2965 CN:  
 CC  
 CC Query Match 100.0%; Score 106; DB 10; Length 26;  
 CC Best Local Similarity 100.0%; Pred. No. 4.81e-05;  
 CC Matches 15: Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 CC  
 CC 1 KSMKVTVAFNSQFQPN 15  
 CC 1 KSMKVTVAFNSQFQPN 15  
 CC  
 CC RESULT 15  
 CC US-08-467-006-#3 STANFARD PAT: 26 AA.  
 CC xxxxxx  
 CC  
 CC Sequence 83: Application US/08467005  
 CC  
 CC Sequence 83: Application US/08467005  
 CC GENERAL INFORMATION:  
 CC APPLICANT: Griffiths, Irwin J.;  
 CC APPLICANT: Pollock, Joanne;  
 CC APPLICANT: Bond, Julian F.;  
 CC APPLICANT: Garman, Richard D;  
 CC APPLICANT: Kuo, Wei-Chang;  
 CC APPLICANT: Yeung, Siu-mei H.;  
 CC APPLICANT: Brauer, Andrew;  
 CC APPLICANT: Exley, Mark A.;  
 CC APPLICANT: Powers, Steven P.  
 CC TITLE OF INVENTION: Allergenic Proteins And Peptides From  
 CC TITLE OF INVENTION: Japanese Cedar Pollen  
 CC NUMBER OF SEQUENCES: 261  
 CC CORRESPONDENCE ADDRESS:  
 CC ADDRESSEE: Immunologic Pharmaceutical Corporation, Inc.  
 CC STREET: 610 Lincoln St  
 CC CITY: Waltham  
 CC STATE: MA  
 CC COUNTRY: USA  
 CC ZIP: 02154

CC COMPUTER READABLE FORM:  
 CC MEDIUM TYPE: Floppy disk  
 CC COMPUTER: IBM PC compatible  
 CC OPERATING SYSTEM: PC-DOS/MS-DOS  
 CC SOFTWARE: Patent Release #1.0, Version #1.25  
 CC CURRENT APPLICATION DATA:  
 CC APPLICATION NUMBER: US/08/467,906  
 CC FILING DATE: June 6, 1995  
 CC CLASSIFICATION: 424  
 CC PRIOR APPLICATION DATA:  
 CC APPLICATION NUMBER: 08/350,225  
 CC FILING DATE: December 6, 1994  
 CC ATTORNEY/AGENT INFORMATION:  
 CC NAME: Jane E. Remillard  
 CC REGISTRATION NUMBER: 38,872  
 CC REFERENCE/DOCKET NUMBER: 025.6 USD4 (IMI-028CPD4);  
 CC TELECOMMUNICATION INFORMATION:  
 CC TELEPHONE: (617) 227-7400  
 CC TELEFAX: (617) 227-5941  
 CC INFORMATION FOR SEQ ID NO: 83:  
 CC SEQUENCE CHARACTERISTICS:  
 CC LENGTH: 26 amino acids  
 CC TYPE: amino acid  
 CC TOPOLOGY: linear  
 CC MOLECULE TYPE: peptide  
 CC FRAGMENT TYPE: internal  
 CC SEQUENCE 26 AA: 2944 MW: 2965 CN:  
 CC  
 CC Query Match 100.0%; Score 106; DB 10; Length 26;  
 CC Best Local Similarity 100.0%; Pred. No. 4.81e-05;  
 CC Matches 15: Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 CC  
 CC 1 KSMKVTVAFNSQFQPN 15  
 CC 1 KSMKVTVAFNSQFQPN 15  
 CC  
 CC Search completed: Mon Jun 19 16:26:18 2000  
 CC Job time : 17 secs.

\*\*\*\*\*  
[W][O][R][D] [P][R][O] [T][E][I][N] [S][E][Q][U][E][N][C][E] [A][N][A][L][Y][S][I][S] [S][O][F][T][W][A][R][E] [V][E][R][S][I][O][N] [1.0] [B][Y] [J][O][H][N] [F.] [C][O][L][L][I][N][S] [B][I][O][C][O][M][P][U][T][I][N][G] [R][E][S][E][A][R][C][H] [U][N][I][T].  
[C][O][P][Y][R][I][G][H][T] (c) 1993-1998 [U][N][I][V][E][R][S][I][T][Y] [O][F] [E][D][I][N][B][U][R][G][H], [U][K].  
[D][I][S][T][R][I][B][U][T][I][O][N] [R][I][G][H][T][S] [B][Y] [O][X][F][O][R][D] [M][O][L][E][C][U][L][A][R] [L][I][D]  
(TM)  
\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd  
MSEARCH\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Mon Jun 19 16:09:54 2000: MasPar time 5.32 Seconds  
Tabular output not generated. 132.94 Million cell updates/sec

Title: >US-09-142-524A-9  
Description: (1-15) from US09.42524A.pap  
Perfect Score: 106  
Sequence: 1 KSMKVTVAFNQFGPN 15  
Scoring table: PAM 150  
Gap 15  
Searched: 142080 seqs, 47172406 residues  
Post-Processing: Minimum Match 0%  
Listing first 45 summaries  
Database: pir62  
1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 26.468 Variance 33.820 scale 0.783  
Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

| SUMMARIES  |       |  |           |
|------------|-------|--|-----------|
| Result No. | Score | Query Match Length DB ID Description               | Pred. No. |
| 1          | 106   | 100.0 374 2 JC2124 major allergen Cry j 1 114e-11  |           |
| 2          | 106   | 100.0 374 2 JC2123 major allergen Cry j 1 114e-11  |           |
| 3          | 80    | 75.5 404 2 S12209 pectate lyase (EC 4.2 3.38e-05   |           |
| 4          | 78    | 73.6 404 2 T05556 pectate lyase (EC 4.2 9.92e-05   |           |
| 5          | 76    | 71.7 455 2 T00356 pectate lyase (EC 4.2 2.88e-04   |           |
| 6          | 74    | 69.8 397 2 S26211 pectate lyase (EC 4.2 8.24e-04   |           |
| 7          | 74    | 69.8 450 2 T09424 probable pectate lyase 8.24e-04  |           |
| 8          | 72    | 67.9 398 2 T07058 pectate lyase (EC 4.2 2.33e-03   |           |
| 9          | 72    | 67.9 434 2 S29612 pectate lyase (EC 4.2 2.33e-03   |           |
| 10         | 72    | 67.9 438 2 S43335 pectate lyase (EC 4.2 2.33e-03   |           |
| 11         | 71    | 67.0 542 2 T06728 pectate lyase (EC 4.2 3.89e-03   |           |
| 12         | 70    | 66.0 418 2 T07701 pectate lyase (EC 4.2 6.48e-03   |           |
| 13         | 68    | 64.2 374 2 T05240 pectate lyase (EC 4.2 1.27e-02   |           |
| 14         | 64    | 60.4 432 2 T00949 hypothetical protein 1.27e-01    |           |
| 15         | 64    | 60.4 449 2 S27098 pectate lyase (EC 4.2 1.27e-01   |           |
| 16         | 64    | 60.4 461 2 T01825 hypothetical protein 1.27e-01    |           |
| 17         | 59    | 55.7 416 2 S45484 glycerolaldehyde-3-phos 1.35e+00 |           |
| 18         | 58    | 54.7 435 1 S31328 omnipotent suppressor 1.35e+00   |           |
| 19         | 57    | 53.8 340 2 G69628 glycerolaldehyde-3-phos 3.35e+00 |           |
| 20         | 57    | 53.8 397 2 C39099 allergen Amb a 1.3 3.35e+00      |           |
| 21         | 57    | 53.8 397 2 C53240 allergen Amb a 1.3 pr 3.35e+00   |           |
| 22         | 57    | 53.8 414 2 S43340 glycerolaldehyde-3-phos 3.35e+00 |           |
| 23         | 57    | 53.8 571 2 E64903 arylsulfatase homolog 3.35e+00   |           |

24 55 51.9 258 1 WMBPP9 lysozyme (EC 3.2.1.17 8.14e+00  
25 55 51.9 258 1 WMBP13 lysozyme (EC 3.2.1.17 8.14e+00  
26 55 51.9 350 2 G71678 multidrug resistance 8.14e+00  
27 55 51.9 398 2 B39099 allergen Amb a 1.2 8.14e+00  
28 55 51.9 398 2 B52240 allergen Amb a 1.2 pr 8.14e+00  
29 54 50.9 138 2 S34735 beta-galactosidase (E 1.2e+01  
30 54 50.9 286 2 T09704 probable arginine/ser 1.2e+01  
31 54 50.9 348 2 S77104 anthranilate phosphor 1.4e+01  
32 54 50.9 355 2 T14013 NADH dehydrogenase su 1.2e+01  
33 54 50.9 355 2 T14008 NADH dehydrogenase su 1.2e+01  
34 54 50.9 396 2 A39099 allergen Amb a 1.1 pr 1.2e+01  
35 53 50.0 340 2 G94666 probable glycerolaldeh 1.93e+01  
36 53 50.0 391 2 T74724 aspartate transaminase 1.93e+01  
37 53 50.0 392 2 C53240 allergen Amb a 1.4 pr 1.93e+01  
38 53 50.0 437 1 S50853 translation releasing 1.93e+01  
39 53 50.0 437 1 A48061 translation releasing 1.93e+01  
40 53 50.0 545 2 JC2500 exo-alpha-sialidase ( 1.93e+01  
41 53 50.0 551 2 C71606 ATP synthase alpha ch 1.93e+01  
42 53 50.0 566 2 G44928 probable acid-cox ii 1.93e+01  
43 53 50.0 599 2 S11361 hypothetical protein 1.93e+01  
44 52 49.1 336 2 T4766 hepatoma transmembran 2.94e+01  
45 52 49.1 355 2 T14310 NADH dehydrogenase su 2.94e+01

ALIGNMENTS

RESULT 1  
ENTRY JC2124 #type complete  
TITLE major allergen Cry j 1 precursor (clone pCC1-15) - Japanese cedar  
ORGANISM #formal\_name Cryptomeria japonica #common\_name Japanese cedar  
DATE 14-Jul-1994 #sequence\_revision 14-Jul-1994 #text\_change 25-Aug-1999  
ACCESSIONS JC2124  
REFERENCE JC2123  
#authors Sone, T.; Koriyama, N.; Shimizu, K.; Kusakabe, T.; Morikubo, K.; Kino, K.  
#journal Biochem. Biophys. Res. Commun. (1994) 199:619-625  
#title Cloning and sequencing of cDNA coding for Cry j 1, a major allergen of Japanese cedar pollen.  
#cross-references MIM:94183234  
#accession JC2124  
#molecule\_type mRNA  
#residues 1-374 #date, SCN  
#cross-references G13265453; N11044663; F11044663; F11044664  
#experimental\_source pollen  
#note the authors described carbohydrate limited site 111 residue 279

CLASSIFICATION #superfamily pectate lyase LA59  
KEYWORDS glycoprotein; pollen  
FEATURE 1-21  
22-374 #domain signal sequence #status predicted #label SIGN  
#product major allergen Cry j 1 (clone pCC1-15) #status predicted #label MATN  
159,191,293,354 #binding\_site carbohydrate (Asn) (covalent) #status predicted  
SUMMARY #length 374 #molecular-weight 40702 #checksum 3692  
Query Match 100.0% Score 106; DB 2; Length 374;  
Best Local Similarity 100.0% Pred. No. 114e-11;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 232 KSMKVTVAFNQFGPN 246  
QY 1 KSMKVTVAFNQFGPN 15  
RESULT 2  
ENTRY JC2123 #type complete  
ENTRY major allergen Cry j 1 precursor (clone pCC1-2-2) - Japanese cedar  
TITLE #formal\_name Cryptomeria japonica #common\_name Japanese cedar  
ORGANISM 14-Jul-1994 #sequence\_revision 14-Jul-1994 #text\_change

```

ACCESSIONS      26-Aug-1999
JC2123: PC2065
REFERENCE
JC2123
#authors      Sano, T.; Komiya, N.; Shimizu, K.; Kusakabe, T.; Morikubo,
              K.; Kino, K.
#journal      Biochem. Biophys. Res. Commun. (1994) 199:619-625
#title        Cloning and sequencing of cDNA coding for Cry j I, a major
              allergen of Japanese cedar pollen.
#cross-references: MIM:3418334
#accession    JC2123
#molecule_type mRNA
#residues     1-374 #label SON
#cross-references: GS:D25544; NID:g493631; PID:d106086; PID:g493632
#experimental_source pollen
#accession    PC2065
#molecule_type protein
#residues     22-53758-81:219-232:236-259:299-307:346-372 #label SO2
#note         the authors described carbohydrate binding site for
              residue 279
#CLASSIFICATION #superfamily pectate lyase LAF59
#KEYWORDS      glycoprotein; pollen
#FEATURE
22-374      #domain: signal sequence #status: predicted #label SIG
#product: major allergen Cry j I (clone PCOI-2-2) #status
              predicted #label MAT
158-191,293,354 #binding-site carbohydrate (Asn) (covalent) #status
              predicted
#SUMMARY      #length 374 #molecular-weight 40645 #checksum 2920

Query Match      100.0%; Score 106; DB 2; Length 374;
Best Local Similarity 100.0%; Pred. No. 1,14e-11;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 232 KSMKVTVAFNQFNP 246
|||||||
QY 1 KSMKVTVAFNQFNP 15

RESULT 3
ENTRY #type complete
#label BUD
#title pectate lyase (EC 4.2.2.2) - tomato
#formal_name Lycopersicon esculentum #common_name tomato
#date 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change
23-Jul-1999
ACCESSIONS      S12209
REFERENCE
S12209
#authors      Roudsley, K.A.; Smith, A.G.; Gasser, S.M.
#journal      Mol. Gen. Genet. (1990) 224:183-192
#title        Regulation of a stylar transmission tissue-specific gene in
              wild-type and transgenic tomato and tobacco.
#cross-references: MIM:9117185
#accession    S12209
#molecule_type mRNA
#residues     1-404 #label BUD
#cross-references: GB:X55193; NID:g19161; PID:CAA38979.1; PID:g19162
#CLASSIFICATION #superfamily pectate lyase LAF59
#KEYWORDS      carbon-oxygen lyase
#SUMMARY      #length 404 #molecular-weight 44298 #checksum 4917

Query Match      75.5%; Score 80; DB 2; Length 404;
Best Local Similarity 76.9%; Pred. No. 3.38e-05;
Matches 10; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

DB 262 KKMQVTIAFNHFG 274
|||||||
QY 1 KSMKVTVAFNQFG 13

RESULT 4
ENTRY #type complete
#label BUD
#title pectate lyase (EC 4.2.2.2) F22K18.20 - Arabidopsis thaliana
              protein F22K18.20

```

```

ORGANISM      #formal_name Arabidopsis thaliana #common_name mossesed;
              #cross-references: EMBL:AL035556
DATE          23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change
23-Jul-1999
ACCESSIONS      T05556
REFERENCE
T05556
#authors      Bevan, M.; Medler, H.; Medler, E.; Wambutt, R.; Hohn, M.;
              Mewes, H.W.; Mayer, K.F.X.; Schell, J.
#submission    Submitted to the Protein Sequence Database, February 1997
#accession     T05556
#molecule_type DNA
#residues      1-404 #label BEV
#cross-references: EMBL:AL035556
#experimental_source cultivar Columbia; BAC clone F22K18
GENETICS
#map_position  37/2; 267/2
#introns       F22K18.20
#note          #superfamily pectate lyase LAF59
#CLASSIFICATION #superfamily pectate lyase LAF59
#KEYWORDS      carbon-oxygen lyase
#SUMMARY      #length 404 #molecular-weight 44495 #checksum 5547

Query Match      73.6%; Score 78; DB 2; Length 404;
Best Local Similarity 69.2%; Pred. No. 9.92e-05;
Matches 9; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

DB 266 KKMQVTIAFNHFG 278
|||||||
QY 1 KSMKVTVAFNQFG 13

RESULT 5
ENTRY #type complete
#label BUD
#title pectate lyase (EC 4.2.2.2) T0556.14 - Arabidopsis thaliana
#formal_name Arabidopsis thaliana #common_name mossesed;
#date 12-Feb-1999 #seq #revision 12-Feb-1999 #text_change
29-Sep-1999
ACCESSIONS      T00856
REFERENCE
T00856
#authors      Roudsley, S.D.; ... X.; Ketchum, K.A.; Crosby, M.;
              Brandon, R.C.; Sykes, S.M.; Kaul, S.; Mason, J.M.;
              Kertavagis, A.R.; Adams, M.D.; Somerville, C.R.; Venter,
              J.C.
#submission    Submitted to the EMBL Data Library, March 1978
#description    Arabidopsis thaliana cell suspension BAC 12.16, genome
              sequence.
#accession     T00856
#status        translated from GB/EMBL/DDBJ
#molecule_type DNA
#residues      1-455 #label ROC
#cross-references: EMBL:AC002521; NID:g247056; PID:AAC05150.1;
              PID:g2947069
#experimental_source cultivar Columbia
GENETICS
#map_position  2
#introns       66/2; 295/3; 376/3
#note          T0085.14
#CLASSIFICATION #superfamily pectate lyase LAF59
#KEYWORDS      carbon-oxygen lyase
#SUMMARY      #length 455 #molecular-weight 51257 #checksum 611

Query Match      71.7%; Score 76; DB 2; Length 455;
Best Local Similarity 69.2%; Pred. No. 2.88e-04;
Matches 9; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

DB 309 KKMQVTIAFNHFG 321
|||||||
QY 1 KSMKVTVAFNQFG 13

RESULT 6
ENTRY #type complete
#label BUD
#title pectate lyase (EC 4.2.2.2) F22K18.20 - Arabidopsis thaliana

```

```

TITLE      pectate lyase (EC 4.2.2.2) - common tobacco
ORGANISM   #formal_name Nicotiana tabacum #common_name common tobacco
DATE       #sequence_revision 12-Feb-1993 #text_change
29-Sep-1999
ACCESSIONS S26211; S26212; S21933; S22753; S22754
REFERENCE   S26211;
#authors    Rogers, H.J.; Harvey, A.; Lonsdale, D.M.
#journal    Plant Mol. Biol. (1992) 20:493-502
#title      Isolation and characterization of a tobacco gene with
            homology to pectate lyase which is specifically expressed
            during microsporogenesis.
#cross-references EMBL:59343039
#accession   S26211
#molecule_type DNA
#residues    1-397 #label ROC
#cross-references EMBL:X67158; NID:q19907; PIDN:CAA47630.1; PID:q19908
#accession   S26212
#molecule_type mRNA
#residues    119-155; 157-188, 190, 'G', 191-199, 'D', 200, 'R', 203-248,
            'N', 250-381 #label ROC
#cross-references EMBL:X67159
#note        translation of the nucleotide sequence is not complete
REFERENCE    S21933
#authors     Lonsdale, D.M.
#submitted   Submitted to the EMBL Data Library, July 1991
#accession   S21933
#molecule_type DNA
#residues    1-397 #label LON
#cross-references EMBL:X61102; NID:q19981; PIDN:CAA43414.1; PID:q19982
GENETICS
#introns     192/1; 293/2
#classification #superfamily pectate lyase LAT59
#keywords     carbon-oxygen lyase
#summary      #length 397 #molecular-weight 4435; #checksum 1480
Query Match      69.8%; Score 74; DB 2; Length 397;
Best Local Similarity 61.5%; Pred. No. 8,244-04;
Matches          8; Conservative 4; Mismatches 1; Indels 0; Gaps 0;
DB 254 KMKVTVAFNFG 266
1 111111111
QY 1 KSMKVTVAFNFG 13
RESULT 7
ENTRY      10524 #type complete
TITLE      Probable pectate lyase (EC 4.2.2.2) - alfalfa
ORGANISM   #formal_name Medicago sativa #common_name alfalfa
DATE       #sequence_revision 16-Jul-1999 #text_change
26-Aug-1999
ACCESSIONS T09524
REFERENCE   Z16712
#authors    Wu, Y.; Qiu, X.; Du, S.; Erickson, L.
#submitted   Submitted to the EMBL Data Library, November 1995
#accession   T09524
#status     preliminary; translated from GB/EMBL/DBJ
#molecule_type DNA
#residues    1-450 #label WUY
#cross-references EMBL:X41472; NID:q1171160; PID:q1171161
GENETICS
#introns     65/2; 245/3
#description catalyzes cleavage of pectate to oligosaccharides
#classification #superfamily pectate lyase LAT59
#keywords     carbon-oxygen lyase
#summary      #length 450 #molecular-weight 50155 #checksum 3091
Query Match      69.8%; Score 74; DB 2; Length 450;
Best Local Similarity 69.2%; Pred. No. 8,244-04;
Matches          9; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
DB 309 KMQITVAFNFG 321
1 111111111

```

```

QY 1 KSMKVTVAFNFG 13
RESULT 8
ENTRY      T07058 #type complete
TITLE      Pectate lyase (EC 4.2.2.2) LAT56 - tomato
ALTERNATE_NAMES protein p56
ORGANISM   #formal_name Lycopersicon esculentum #common_name tomato
DATE       #sequence_revision 30-Apr-1999 #text_change
29-Sep-1999
ACCESSIONS T07058; S08383
REFERENCE   Z15890
#authors    Wang, R.A.
#submitted   Submitted to the EMBL Data Library, September 1994
#accession   T07058
#status     translated from GB/EMBL/DBJ
#molecule_type DNA
#residues    1-398 #label WIN
#cross-references EMBL:X15130; NID:q551555; PIDN:CAA31524.1;
            #cross-references PID:q55116
#experimental_source strain: VF35; author
            S08383
#authors     Wang, R.A.; Yamauchi, C.; Databell, S.K.; Ursula, V.K.;
            McCormick, S.
#journal     Plant Mol. Biol. (89) 14:17-28
#title       Molecular and genetic characterization of two
            pollen-expressed genes that have sequence similarity to
            pectate lyases (the plant pathogen Erwinia
            #cross-references EMBL:91322485
#accession   S08383
#molecule_type DNA
#residues    1-24; 126; 131-61; 63-102; 104-166; 167-194; 195-
            118-141; 142-143-156; SNSQYCYRSL177-349; MVCH
            354-355; 357 #label KIW
#cross-references EMBL:X15500
GENETICS
#map_position 194/1; 234/2
#introns      #superfamily pectate lyase LAT59
#classification carbon-oxygen lyase
#keywords      #length 338 #molecular-weight 44563 #checksum 5206
Query Match      57.9%; Score 72; DB 2; Length 338;
Best Local Similarity 53.8%; Pred. No. 2,406-03;
Matches          7; Conservative 6; Mismatches 0; Indels 0; Gaps 0;
DB 455 KMKVTVAFNFG 49
1 111111111
QY 1 KSMKVTVAFNFG 13
RESULT 9
ENTRY      S29612 #type complete
TITLE      Pectate lyase (EC 4.2.2.2) - trumpet lily
ORGANISM   #formal_name Lilium longiflorum #common_name trumpet lily
DATE       #sequence_revision 19-Mar-1997 #text_change
29-Sep-1999
ACCESSIONS S29612
REFERENCE   S29611
#authors     Kim, S.; Finkbeiner, D.J.; An, G.
#submitted   Submitted to the EMBL Data Library, October 1992
#description Abundancy patterns of lily pollen cDNAs: characterization of
            three pollen-preferential cDNA clones.
#accession   S29612
#status     preliminary
#molecule_type mRNA
#residues    1-434 #label KIM
#cross-references EMBL:217328; NID:q19450; PIDN:CAA78976.1; PID:q19451
#experimental_source cv. Nellie White, mature flower
#classification #superfamily pectate lyase LAT59
#keywords      carbon-oxygen lyase
#summary      #length 434 #molecular-weight 48457 #checksum 6432

```

Query Match 67.9% Score 72: DB 2: Length 434:  
 Best Local Similarity 81.8% Pred. No. 2.33e-03:  
 Matches 9: Conservative 1: Mismatches 0: Indels 0: Gaps 0:

DB 296 MKQVIAFNHFG 306  
 1 1 1 1 1 1 1 1 1 1  
 QY 3 MKQVIAFNHFG 13

RESULT 10 S43335 #type complete  
 ENTRY pectate lyase (EC 4.2.2.2) - maize  
 TITLE #formal\_name zea mays #common\_name maize  
 ORGANISM 07-Sep-1994 #sequence\_revision 10-Nov-1995 #text\_change  
 DATE 29-Sep-1999  
 ACCESSIONS S43335  
 REFERENCE S43334  
 #authors Turcich, M.P.; Hamilton, D.A.; Mascarenhas, J.P.  
 #journal Plant Mol. Biol. (1999) 23:1061-1065  
 #title Isolation and characterization of pollen-specific maize genes  
 with sequence homology to radweed allergens and pectate  
 lyases.

##cross-references: MIM:9403558  
 #accession S43335  
 ##status nucleic acid sequence not shown  
 ##molecule\_type DNA  
 ##residues 1-438 ##label TUR  
 ##cross-references EMBL:L26140; NID:9405534; PDB:AAA16476.1;  
 PDB:9405535  
 CLASSIFICATION #superfamily pectate lyase LAT59  
 KEYWORDS carbon-oxygen lyase  
 SUMMARY #length 438 #molecular-weight 49138 #checksum 1084

Query Match 67.9% Score 72: DB 2: Length 438:  
 Best Local Similarity 81.8% Pred. No. 2.33e-03:  
 Matches 9: Conservative 1: Mismatches 1: Indels 0: Gaps 0:

DB 300 MKQVIAFNHFG 510  
 1 1 1 1 1 1 1 1 1 1  
 QY 3 MKQVIAFNHFG 13

RESULT 11 106724 #type complete  
 ENTRY pectate lyase (EC 4.2.2.2) F28P10 Arabidopsis thaliana  
 TITLE Protein F28P10.100  
 ALTERNATE\_NAMES #formal\_name Arabidopsis thaliana #common\_name mouse ear  
 ORGANISM cress

DATE 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change  
 DATE 23-Jul-1999

ACCESSIONS 106728  
 REFERENCE 215793  
 #authors Quetier, F.; Choisme, N.; Robert, C.; Brettiere, P.; Wincker,  
 P.; Cattolico, L.; Artiguenave, F.; Saurin, W.;  
 Weissbach, J.; Salanoubat, M.; Mewes, H.W.; Mayer,  
 K.F.X.; Schaefer, C.

#submission submitted to the Protein Sequence Database, April 1999  
 #accession 106728  
 ##molecule\_type DNA

##residues 1-542 ##label QYE  
 ##cross-references EMBL:AL049655  
 ##experimental\_source cultivar Columbia: BAC clone F28P10

GENETICS  
 #map\_position 1  
 #introns 46/2: 346/3: 413/2: 480/2  
 #note F28P10.100  
 CLASSIFICATION #superfamily pectate lyase LAT59  
 KEYWORDS carbon-oxygen lyase  
 SUMMARY #length 542 #molecular-weight 58573 #checksum 1032

Query Match 67.0% Score 71: DB 2: Length 542:  
 Best Local Similarity 66.7% Pred. No. 3.89e-03:  
 Matches 8: Conservative 3: Mismatches 1: Indels 0: Gaps 0:

DB 280 GMQVIAFNHFG 291  
 1 1 1 1 1 1 1 1 1 1  
 QY 2 SMKVIAFNHFG 13

RESULT 12 107701 #type complete  
 ENTRY pectate lyase (EC 4.2.2.2) F17N18.100 Arabidopsis thaliana  
 TITLE Protein F17N18.100  
 ORGANISM #formal\_name Arabidopsis thaliana #common\_name nullified:  
 cress

DATE 14-May-1999 #sequence\_revision 14-May-1999 #text\_change  
 DATE 23-Jul-1999

ACCESSIONS 107701  
 REFERENCE 215098  
 #authors Bevan, M.; Pohl, T.; Weizenecker, E.; Bamzait, E.; Mewes

#submission submitted to the Protein Sequence Database, May 1999  
 #accession 107701  
 ##molecule\_type DNA

##residues 1-418 ##label BEV  
 ##cross-references EMBL:AL049751  
 ##experimental\_source cultivar Columbia: BAC clone F17N18

GENETICS  
 #map\_position 4  
 #introns 26/2: 49/2: 264/3: 289/2: 345/3  
 #note F17N18.100  
 CLASSIFICATION #superfamily pectate lyase LAT59  
 KEYWORDS carbon-oxygen lyase  
 SUMMARY #length 418 #molecular-weight 46251 #checksum 599

Query Match 66.0% Score 70: DB 2: Length 418:  
 Best Local Similarity 61.5% Pred. No. 6.45e-03:  
 Matches 8: Conservative 3: Mismatches 2: Indels 0: Gaps 0:

DB 278 KMQVIAFNHFG 290  
 1 1 1 1 1 1 1 1 1 1  
 QY 1 KMKVIAFNHFG 13

RESULT 13 105240 #type complete  
 ENTRY pectate lyase (EC 4.2.2.2) F18A5.100 Arabidopsis thaliana  
 TITLE Protein F18A5.100  
 ALTERNATE\_NAMES #formal\_name Arabidopsis thaliana #common\_name mouse ear:  
 cress

DATE 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change  
 DATE 23-Jul-1999

ACCESSIONS 105240  
 REFERENCE 215405  
 #authors Bevan, M.; Weber, N.; Grueninger, D.; Schmidhafer, T.;  
 Bamzait, E.; Mewes, H.W.; Mayer, K.F.X.; Schaefer, C.

#submission submitted to the Protein Sequence Database, February 1999  
 #accession 105240  
 ##molecule\_type DNA

##residues 1-374 ##label BEV  
 ##cross-references EMBL:AL035528  
 ##experimental\_source cultivar Columbia: BAC clone F18A5

GENETICS  
 #map\_position 4  
 #note Intron positions not resolved  
 #note F18A5.100  
 CLASSIFICATION #superfamily pectate lyase LAT59  
 KEYWORDS carbon-oxygen lyase  
 SUMMARY #length 374 #checksum 1983

Query Match 64.2% Score 68: DB 2: Length 374:  
 Best Local Similarity 61.5% Pred. No. 1.78e-02:  
 Matches 8: Conservative 3: Mismatches 2: Indels 0: Gaps 0:

DB 234 KLMQVIAFNHFG 246









FT ACT\_SITE 272 POTENTIAL.  
 FT CARBOHYD 134 POTENTIAL.  
 FT CARBOHYD 227 POTENTIAL.  
 FT CONFLICT 156 S -> C (IN MRNA).  
 FT CONFLICT 189 GS -> SG (IN MRNA).  
 FT CONFLICT 200 S -> D (IN MRNA).  
 FT CONFLICT 202 H -> R (IN MRNA).  
 FT CONFLICT 249 H -> N (IN MRNA).  
 SQ SEQUENCE 397 AA: 44351 MW: EFC82CE5DA7643F CRC64:

Query Match 59.8% Score 74: DB 1: Length 397:  
 Best Local Similarity 61.5% Pred. No. 1:05e-04:  
 Matches 8: Conservative 4: Mismatches 1: Indels 0: Gaps 0:

DB 254 KMKITLAYNHFG 266  
 QY 1 KSMKVTVAFNQFG 13

RESULT 4  
 ID PE55\_LYCES STANDARD: PRT: 398 AA.

AC P15721:  
 DT 01-APR-1990 (Rel. 14, Created)  
 DT 01-FEB-1998 (Rel. 33, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE PROBABLE PECTATE LYASE P56 PRECURSOR (EC 4.2.2.2).  
 GN LAT56.  
 OS Lycopersicon esculentum (Tomato).  
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 CC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;  
 CC core eudicots; Asteridae; euasterids I; Solanales; Solanaceae;  
 CC Solanum.  
 RN [1]

SEQUENCE FROM N.A.  
 P1 STRAIN-CV, VF36; TISSUE=ANTHER;  
 RX MEDLINE: 91322485  
 RA Wang R.A., Yamaguchi J., Larabell S.K., Frisn V.M., McCormick S.;  
 RT "Molecular and genetic characterization of two pollen-expressed genes  
 R1 that have sequence similarity to pectate lyases of the plant pathogen  
 R2 Erwinia";  
 R3 Plant Mol. Biol. 14:17-28(1990).  
 RN [2]

RP REVISIONS.  
 RC STRAIN-CV, VF36; TISSUE=ANTHER;  
 RA Wang R.A.;  
 RL Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.  
 CC FUNCTION: MIGHT BE NEEDED DURING POLLEN DEVELOPMENT. AND TUBE  
 CC GROWTH.  
 CC CATALYTIC ACTIVITY: ELIMINATIVE CLEAVAGE OF PECTATE TO GIVE  
 CC OLIGOSACCHARIDES WITH 4-DEOXY-ALPHA-D-MANN-4-ENURONOSYL GROUPS  
 CC AT THEIR NON-REDUCING ENDS.  
 CC TISSUE SPECIFICITY: EXPRESSED IN ANTHERS AND POLLEN.  
 CC SIMILARITY: BELONGS TO THE POLYSACCHARIDE LYASE FAMILY 1.

This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL: X15500; CAA33524.1;  
 DR PIR: S08383; S08383.  
 DR PFAM: PF00544; pec\_lyase; 1.  
 DR PRINTS: PR00807; AMBALLERGEN.  
 KW Lyase; Multigene family; Signal;  
 FT SIGNAL 1 27 OR 22 (POTENTIAL).  
 FT CHAIN 28 398 PROBABLE PECTATE LYASE P56.  
 FT ACT\_SITE 273 273 POTENTIAL.  
 FT CARBOHYD 135 135 POTENTIAL.  
 FT CARBOHYD 228 228 POTENTIAL.  
 SQ SEQUENCE 398 AA: 44563 MW: 8D676250BDB8C7C8 CRC64:

Query Match 67.9% Score 72: DB 1: Length 398:  
 Best Local Similarity 53.8% Pred. No. 3.34e-04:  
 Matches 7: Conservative 6: Mismatches 0: Indels 0: Gaps 0:

DB 255 RGMKITLAYNHFG 267  
 QY 1 KSMKVTVAFNQFG 13

RESULT 5  
 ID PE5\_LILLO STANDARD: PRT: 434 AA.

AC P40973:  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE PECTATE LYASE PRECURSOR (EC 4.2.2.2).  
 OS Lilium ioniflorum (Trumpet lily).  
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 CC euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Liliales;  
 CC Liliaceae; Lilium.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CV, NELLIE WHITE; TISSUE=POLLEN.  
 RA Kim S.R., Finkel D.J., An G.;  
 RL Submitted (JUN-1993) to the EMBL/GenBank/DBJ databases.  
 CC CATALYTIC ACTIVITY: ELIMINATIVE CLEAVAGE OF PECTATE TO GIVE  
 CC OLIGOSACCHARIDES WITH 4-DEOXY-ALPHA-D-MANN-4-ENURONOSYL GROUPS  
 CC AT THEIR NON-REDUCING ENDS.

CC SIMILARITY: BELONGS TO THE POLYSACCHARIDE LYASE FAMILY 1.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL: Z17328; CAA78976.1;  
 DR EMBL: L18911; AAA33398.1;  
 DR PIR: S29612; S29612.  
 DR PFAM: PF00544; pec\_lyase; 1.  
 DR PRINTS: PR00807; AMBALLERGEN.  
 KW Lyase; Signal;  
 FT SIGNAL 1 24 POTENTIAL.  
 FT CHAIN 23 434 PECTATE LYASE.  
 FT ACT\_SITE 312 312 POTENTIAL.  
 FT CARBOHYD 68 68 POTENTIAL.  
 FT CARBOHYD 97 97 POTENTIAL.  
 SQ SEQUENCE 434 AA: 48457 MW: C1F3E30AD2B5D064 CRC64:

Query Match 67.9% Score 72: DB 1: Length 434:  
 Best Local Similarity 81.8% Pred. No. 3.34e-04:  
 Matches 9: Conservative 1: Mismatches 1: Indels 0: Gaps 0:

DB 296 MGVTVAFNHFG 306  
 QY 3 KMKVTVAFNQFG 13

RESULT 6  
 ID PE59\_LYCES STANDARD: PRT: 449 AA.

AC P15722:  
 DT 01-APR-1990 (Rel. 14, Created)  
 DT 01-APR-1990 (Rel. 14, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE PROBABLE PECTATE LYASE P59 PRECURSOR (EC 4.2.2.2).  
 GN LAT59.  
 OS Lycopersicon esculentum (Tomato).  
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 CC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;  
 CC core eudicots; Asteridae; euasterids I; Solanales; Solanaceae;

RN Solanum.  
 RP [1]  
 RC SEQUENCE FROM N.A.  
 RC STRAIN-CV, VF35; TISSUE-ANTHER;  
 RX MEDLINE: 9322485.  
 RA Wang R.A., Yabuzaki J., Jarabell S.K., Ursin V.M., McCormick S.  
 RI Molecular and genetic characterization of two pollen-expressed genes  
 RI that have sequence similarity to pectate lyases of the plant pathogen  
 RI Erwinia.  
 RL Plant Mol. Biol. 14:17-28(1990).  
 CC [1] FUNCTION: MIGHT BE NEEDED DURING POLLEN DEVELOPMENT AND TUBE  
 CC GROWTH.  
 CC [1] CATALYTIC ACTIVITY: ELIMINATIVE CLEAVAGE OF PECTATE TO GIVE  
 CC OLIGOSACCHARIDES WITH 4-DEOXY-ALPHA-D-MANN-4-EPIMERISYL GROUPS  
 CC AT THEIR NON-REDUCING ENDS.  
 CC [1] TISSUE SPECIFICITY: EXPRESSED IN ANTHERS AND POLLEN.  
 CC [1] SIMILARITY: BELONGS TO THE POLYSACCHARIDE LYASE FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation.  
 CC use by non-profit institutions is allowed as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: X15499; CAA13523.1; .  
 CC PIR: S27098; S27098.  
 CC PFAM: PF00344; PectLyase; 1.  
 CC PRINTS: PR00807; AMBALLERGEN.  
 CC Lyase: Multigene family; Signal.  
 CC SIGNAL 22 POTENTIAL.  
 CC CHAIN 21 449 PROBABLE PECTATE LYASE P59.  
 CC ACT SITE 325 325 POTENTIAL.  
 CC CARBOXYD 56 56 POTENTIAL.  
 CC CARBOXYD 80 80 POTENTIAL.  
 CC CARBOXYD 81 81 POTENTIAL.  
 CC SEQUENCE 449 AA: 50893 MW: 1782AA13F:735C3C CAC64.  
 Query Match 60.4% Score 64 DB 1 Length 449  
 Best Local Similarity 63.6% Pred. No. 2 82e-02.  
 Matches 7: Conservative 3: Mismatches 1: Indels 0: Gaps 0:

DB 309 MOLLUSCAEPP 319  
 CY 3 MKVIVAFN0FG 13  
 RESULT 7  
 AC GSEA GRAVE STANDARD: PFI: 416 AA  
 DI P30724:  
 DI 01-APR-1993 (Rel. 25, Created)  
 DI 01-APR-1993 (Rel. 25, Last sequence update)  
 DI 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE GLYCERALDEHYDE 3-PHOSPHATE DEHYDROGENASE, CHLOROPLAST PRECURSOR  
 DE (EC 1.2.1.12).  
 GN GAPA.  
 OS Gracilaria verrucosa.  
 OC Eukaryota; Rhodophyta; Florideophyceae; Gracilariaceae; Gracilariaceae;  
 OC Gracilaria.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-NORMAN;  
 RX MEDLINE: 93306755.  
 RA Zhou Y.H., Ragun M.A.  
 RI "cDNA cloning and characterization of the nuclear gene encoding  
 RI chloroplast glyceraldehyde-3-phosphate dehydrogenase from the marine  
 RI red alga Gracilaria verrucosa."  
 RL Curr. Genet. 23:483-489(1993).  
 CC [1]  
 CC REVIEWS:  
 CC MEDLINE: 95042840.  
 CC Zhou Y.H., Ragun M.A.

RI glyceraldesdehyde-3-phosphate dehydrogenase from the marine red alga  
 RI Gracilaria verrucosa.  
 RL Curr. Genet. 26:79-86(1994).  
 CC [1] CATALYTIC ACTIVITY: D-GLYCERALDEHYDE 3-PHOSPHATE + NADPH + H<sub>2</sub>O  
 CC + NAD(+) -> 1,3-BIPHOSPHATETETRAATE + NADH.  
 CC [1] PATHWAY: CALVIN CYCLE.  
 CC [1] SUBUNIT: HOMOTETRAMER.  
 CC [1] SUBCELLULAR LOCATION: CHLOROPLAST.  
 CC [1] SIMILARITY: BELONGS TO THE GLYCERALDEHYDE 3-PHOSPHATE  
 CC DEHYDROGENASE FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation.  
 CC use by non-profit institutions is allowed as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: Z15102; CAA78811.1; .  
 CC EMBL: L22011; AAA33355.1; .  
 CC PIR: S25596; S25596.  
 CC PIR: S45484; S45484.  
 CC HSSP: P50362; INH.  
 CC MENDEL: 560; GRAVE; GAPA.1.  
 CC PRINTS: PR00078; G3PDHGNASE.  
 CC PROSITE: PS00071; GAPDH; 1.  
 CC PFAM: PF00044; gpdh; 1.  
 CC Calvin cycle; Oxidoreductase; NAD; Transit peptide; Chloroplast.  
 KW Multigene family.  
 FT CHAIN 1 78 CHLOROPLAST (POTENTIAL).  
 FT CHAIN 79 416 GLYCERALDEHYDE 3-PHOSPHATE DEHYDROGENASE.  
 FT BINDING 233 233 GLYCERALDEHYDE 3-PHOSPHATE.  
 FT ACT SITE 260 260 A 11VATIS TRLG GROUP DURING CALVIN CYCLE.  
 FT SEQUENCE 416 AA: 44337 MW: 43594.7E4230B8 CAC64.  
 Query Match 55.7% Score 59 DB 1 Length 416  
 Best Local Similarity 75.0% Pred. No. 3 83e-01  
 Matches 9: Conservative 1: Mismatches 2: Indels 0: Gaps 0:

DB 75 SMKVIVAFN0FG 89  
 CY 2 SMKVIVAFN0FG 13  
 RESULT 8  
 AC EPI LABATH STANDARD: PFI: 416 AA  
 DI P35614:  
 DI 01-JUN-1994 (Rel. 29, Created)  
 DI 01-JUN-1994 (Rel. 29, Last sequence update)  
 DI 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE EUKARYOTIC PEPTIDE CHAIN RELEASE FACTOR SUBUNIT 1 (ERF1) (CWNIP-1EN)  
 DE SUPPRESSOR PROTEIN 1 HOMOLOG; (SUP1 HOMOLOG).  
 GN ERF1-3.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC eudicotyledons; Spermatophytes; Magnoliopsida; eudicotyledons;  
 OC core eud. cots; Rosidae; eustosids 11; Brassicales; Brassicaceae;  
 OC Arabidopsis.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CV, COLUMBIA;  
 RA Brown C.M., Quigley F.R., Miller W.A.;  
 RI "Three eukaryotic release factor one (erf1) homologs from Arabidopsis  
 RI thaliana Columbia."  
 RL Plant Gene Register PGR95-123.  
 CC [1]  
 CC REVIEWS:  
 CC MEDLINE: 97386699.  
 CC Quigley F., Dac P., Cottet A., Maché R.;  
 CC "Sequence analysis of an 81 kb contig from Arabidopsis thaliana

RI Chromosome III.  
RL Nucleic Acids Res. 24:4312-4318(1996).  
CC - FUNCTION: DIRECTS THE TERMINATION OF TRANSLATION IN RESPONSE TO  
CC ALL THREE PEPTIDE CHAIN TERMINATION CODONS (BY SIMILARITY).  
CC - SUBUNIT: HETERODIMER OF TWO SUBUNITS, ONE OF WHICH BINDS GTP.  
CC - SIMILARITY: EXTREMELY HIGH, TO OTHER SPECIES ENFL.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: X65375; CAA49172.1  
CC EMBL: X98110; CAA66813.1  
CC EMBL: X97486; CAA66218.1  
CC PIR: S31328; S31328.  
CC PFAM: PF01605; RFT: 1.  
CC Protein biosynthesis.  
CC  
CC Query Match 54.7% Score 58: DB 1: Length 435;  
CC Best Local Similarity 50.0% Pred. No. 6.34e-01;  
CC Matches 7: Conservative 3: Mismatches 4: Indels 0: Gaps 0:  
CC  
CC 104 KKKVTDIEFFXP 117  
CC 1 KSKKVIIVAFNFGP 14  
CC  
CC  
CC RESULT 9  
CC ID G3P2\_BACSU STANDARD: PRT: 340 AA.  
CC Q3425:  
CC 15-JUL-1998 (Rel. 36, Created)  
CC 15-JUL-1998 (Rel. 36, Last sequence update)  
CC 15-JUL-1998 (Rel. 36, Last annotation update)  
CC DE GLYCERALDEHYDE 3-PHOSPHATE DEHYDROGENASE 2 (EC 1.2.1.12) (GAPDH).  
CC GABP.  
CC GN  
CC OS Bacteria; subtilis.  
CC CC Bacteria; Firmicutes; Bacillus/Clostridium group;  
CC CC Bacillus/Staphylococcus group; Bacillus.  
CC RN  
CC RF  
CC RE MEDLINE: 98048467.  
CC RA Lapidus A., Gallerton N., Sorokin A., Ehrlich S.D.,  
CC RI in the 200 kb trans-dnaB region.  
CC RL Microbiology 143:3431-3441(1997).  
CC CC - CATALYTIC ACTIVITY: D-GLYCERALDEHYDE 3-PHOSPHATE + ORTHOPHOSPHATE  
CC - NAD(+) = 1,3-DIPHOSPHATE + GLYCERATE + NADH.  
CC - SUBUNIT: HOMOTETRAMER (BY SIMILARITY).  
CC - SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).  
CC - SIMILARITY: BELONGS TO THE GLYCERALDEHYDE 3-PHOSPHATE  
CC DEHYDROGENASE FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: AF008220; AAC00355.1  
CC EMBL: Z99118; CAB14862.1  
CC HSP: P0C362; 3DEV.  
CC SUBTILIST: B012592; GABP.  
CC PRINTS: PR00079; G3PDHGRGNASE.  
CC PROSITE: PS00071; GAPDH: 1.  
CC PFAM: PF00044; gpdh: 1.  
CC Glycolysis: Oxidoreductase; NAD: Multigene family.

FT BINDING 152 152 GLYCERALDEHYDE 3-PHOSPHATE.  
FT ACT\_SITE 179 179 ACTIVATES THIOL GROUP DURING CATALYSIS.  
SQ SEQUENCE 340 AA: DS1DB4BC7D533E26 CRC64:  
Query Match 53.8% Score 57: DB 1: Length 340;  
Best Local Similarity 72.7% Pred. No. 1.04e-00;  
Matches 8: Conservative 1: Mismatches 2: Indels 0: Gaps 0:  
DB 1 MKKVAINGFG 11  
QY 3 MKVIVAFNFGP 13  
RESULT 10  
ID MP13\_AMBAR STANDARD: PRT: 397 AA.  
AC P27761:  
DT 01-AUG-1992 (Rel. 23, Created)  
DT 01-AUG-1992 (Rel. 23, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE POLLEN ALLERGEN AMB A 1.3 PRECURSOR (ANTIGEN E) (ANTIGEN AMB A 1)  
OS Ambrosia artemisiifolia (Short ragweed).  
CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
CC Euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;  
CC core eudicots; Asterales; easterids II; Asterales; Asteraceae;  
CC Asteroidae; Helianthaceae; Ambrosia.  
CC RN  
CC RP SEQUENCE FROM N.A.  
CC RC  
CC RE TISSUE: POLLEN;  
CC RA Rafnar T., Griffith I.J., Kuo M.-C., Bond J.F., Rogers B.L.,  
CC RI Klapper D.G.,  
CC RT Cloning of Amb a I (antigen E), the major allergen family of short  
CC ragweed pollen.  
CC RL J. Biol. Chem. 265:1229-1236(1991).  
CC RN [2]  
CC RP SEQUENCE FROM N.A., AND V. ANTIS.  
CC RC  
CC RE TISSUE: POLLEN;  
CC RA Griffith I.J., Pollock J., Klapper D.G., Rogers B.L., Nault A.M.,  
CC RI "Sequence polymorphism of Amb a I and Amb a II, the major allergens  
CC in Ambrosia artemisiifolia (short ragweed)."  
CC RL Int. Arch. Allergy Appl. Immunol. 96:296-304(1991).  
CC CC - SUBUNIT: MONOMER.  
CC CC - TISSUE SPECIFICITY: POLLEN AND FLOWERS.  
CC CC - DISEASE: ONE OF THE MAJOR ALLERGENS OF THE RAGWEED POLLEN.  
CC CC - SIMILARITY: BELONGS TO THE GLYCERALDEHYDE 3-PHOSPHATE  
CC AMB A I/AMB A II/CRY 3: SUBFAMILY  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: M62961; AAA32668.1  
CC EMBL: M80560; AAA32669.1; ALT-SEQ.  
CC PIR: C39099; C39099.  
CC PFAM: PF00544; pec\_lyase: 1.  
CC PRINTS: PR00807; AMBALLERGEN.  
CC Antigen: Allergen: Signal; Multigene family: Polymorphism.  
FT SIGNAL 1 25 POTENTIAL.  
FT CHAIN 26 397 POLLEN ALLERGEN AMB A 1.3.  
FT MOD\_RES 26 397 BLOCKED.  
FT VARIANT 48 48 L -> Y.  
SQ SEQUENCE 397 AA: 42928 MW: C8DB4:257590DD0A CRC64:  
Query Match 53.8% Score 57: DB 1: Length 397;  
Best Local Similarity 50.0% Pred. No. 1.04e-00;  
Matches 9: Conservative 1: Mismatches 5: Indels 0: Gaps 0:  
DB 256 KGLIATVAFNMFDTN 270

1 KSMKVTVAFCNFGP 15

RESULT 13  
ID G3PA\_CHOCR STANDARD: PRT: 414 AA.

AC P34919:  
D: 01-FEB-1994 (Rel. 28, Created)  
D: 01-FEB-1994 (Rel. 28, Last sequence update)  
D: 15-JUL-1998 (Rel. 38, Last annotation update)  
D: 15-JUL-1998 (Rel. 38, Last annotation update)  
D: GLYCERALDEHYDE 3-PHOSPHATE DEHYDROGENASE, CHLOROPLAST PRECURSOR  
D: (EC 1.2.1.12)  
D: GAP.

OS Chondrus crispus (Carrageen).  
OS Eukaryota: Rhodophyta: Floridophyceae: Gigartinales: Gigartiniaceae:  
OS Chondrus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 94275876.

RA Liud M.-E., Valentin C., Martin W., Bouget F.Y., Kieareg B.,  
RA Cerff R.  
RA "The evolutionary origin of red algae as deduced from the nuclear  
RA genes encoding chlorophyll and chloroplast glyceraldehyde-3-phosphate  
RA dehydrogenases from Chondrus crispus."  
RA J. Mol. Evol. 38:319-327(1994).

CC -1- CATALYTIC ACTIVITY: D-GLYCERALDEHYDE 3-PHOSPHATE + ORTHOPHOSPHATE  
CC + NAD(+) -> 1,3-DIPHOSPHATEGLYCERATE + NADH.  
CC -1- PATHWAY: CALVIN CYCLE.  
CC -1- SUBUNIT: HOMOTETRAMER.  
CC -1- SUBCELLULAR LOCATION: CHLOROPLAST.  
CC -1- SIMILARITY: BELONGS TO THE GLYCERALDEHYDE 3-PHOSPHATE  
CC DEHYDROGENASE FAMILY.

CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation.  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

EMBL: X73035; CAA51515.1;  
D: PIR: S32705; S32705.  
D: HSSP: P50362; INH.  
D: MENDEL: 7723; CHRCAT:CaAl1.  
D: PRINTS: P00078; GPPDHORINASE.  
D: PROSITE: PS00071; GAPDH.  
D: PFAM: PFG044; qpdh.1.  
KW Calvin cycle; oxidoreductase; NADH; NADP; NADPH; NADP+; NADPH+;  
KW Multisubunit family.  
FT TRANSIT 76 CHLOROPLAST (POTENTIAL);  
FT CHAIN 77 414 GLYCERALDEHYDE 3-PHOSPHATE DEHYDROGENASE.  
FT BINDING 231 231 GLYCERALDEHYDE 3-PHOSPHATE.  
FT ACT\_SITE 258 258 ACTIVATES THIOL GROUP DURING CATALYSIS.  
SQ SEQUENCE 414 AA: 44459 MW: 96505262C61064111 CRC64:

Query Match 53.8% Score 57: DB 1: Length 414;  
Best Local Similarity 66.7% Pred. NO. 1:04e-00;  
Matches 8: Conservative 2: Mismatches 2: Indels 0: Gaps 0;

13 75 TMKRVYVAINCFG 87  
14 2 SKMKVTVAFCNFG 13

RESULT 12  
ID LYCV\_BP2H2 STANDARD: PRT: 258 AA.

AC P11187:  
D: 01-JUL-1989 (Rel. 11, Created)  
D: 01-JUL-1989 (Rel. 11, Last sequence update)  
D: 01-NOV-1997 (Rel. 35, Last annotation update)  
D: LYSOZYME (EC 3.2.1.17) (LYSIS PROTEIN) (MURAMIDASE) (MORPHOGENESIS)

DE GN  
OS Bacteriophage phi-29.  
OC Viruses: dsDNA viruses, no RNA stage: Tailed phages: Podoviridae.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 87117505.  
RA Garvey K.J., Saedi M.S., Ito J.  
RA "Nucleotide sequence of Bacillus phage phi 29 genes 14 and 15:  
RA homology of gene 15 with other phage lysozymes."  
RA Nucleic Acids Res. 14:1000-1008(1986).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 87106857.  
RA Vicek C., Paces V.  
RA "Nucleotide sequence of the late region of Bacillus phage phi 29  
RA completes the 5,285-bp sequence of phi 29 genome. Comparison with  
RA the homologous sequence of phage PZA."  
RA Gene 46:215-225(1986).  
RN [3]  
RP SEQUENCE OF 115-258 FROM N.A.  
RX MEDLINE: 8616973.  
RA Garvey K.J., Saedi M.S., Ito J.  
RA "The complete sequence of Bacillus phage phi 29 gene 16: a primary  
RA required for the genome encapsidation reaction."  
RA Gene 40:311-316(1985).  
CC -1- FUNCTION: HELPS TO RELEASE THE MATURE PARTICLES FROM THE  
CC CELL WALL BY BREAKING DOWN THE PEPTIDOGLYCAN.  
CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF THE 1,4-BETA-LINKAGES BETWEEN  
CC N-ACETYL-D-GLUCOSAMINE AND N-ACETYL-MURAMIC ACID IN PEPTIDOGLYCAN  
CC HETEROPOLYMERS OF THE PROKARYOTES CELL WALLS.  
CC -1- SIMILARITY: BELONGS TO FAMILY 24 OF GLYCOSYL HYDROLASES  
CC THIS SWISS-PROT entry is copyrighted. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation.  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

EMBL: X04962; CAA28632.1;  
D: ENBL: M14782; AAA22286.1;  
D: EMSC: M14431; AAA8347.1;  
D: PIR: B24721; WKHP9.  
D: HSSP: P07231; Z05L.  
D: PFAM: PFC0959; Phage\_lysozyme.  
D: PFAM: PFC1476; P3\_Bacillus\_2.  
KW Late protein; hydrolyses glycosidic bonds; bacteriolytic enzyme  
FT ACT\_SITE 15 15 PROTON DONOR (BY SIMILARITY)  
SQ SEQUENCE 258 AA: 28094 MW: 264500192784208 CRC64:

Query Match 51.9% Score 55: DB 1: Length 258;  
Best Local Similarity 50.8% Pred. NO. 2:77e-00;  
Matches 7: Conservative 5: Mismatches 2: Indels 0: Gaps 0;

DB 70 KALKVSVTONCFDA 83  
QY 1 KSMKVTVAFCNFGP 14

RESULT 13  
ID LYCV\_BP2H2 STANDARD: PRT: 258 AA.  
AC P07540:  
D: 01-APR-1988 (Rel. 07, Created)  
D: 01-APR-1988 (Rel. 07, Last sequence update)  
D: 01-FEB-1996 (Rel. 33, Last annotation update)  
D: LYSOZYME (EC 3.2.1.17) (LYSIS PROTEIN) (MURAMIDASE) (LATE PROTEIN)  
D: GP15).  
D: 15.  
D: Bacteriophage PZA.  
OS Viruses: dsDNA viruses, no RNA stage: Tailed phages: Podoviridae.  
RN [1]

RP SEQUENCE FROM N.A.  
 PA MEDLINE: 87031573.  
 RA Paces V., Vacek C., Urbanek P.:  
 RI "nucleotide sequence of the late region of *Bacillus subtilis* phage  
 PZA, a close relative of phi 29.";  
 RL Gene 44:107-114(1986).  
 CC -1- FUNCTION: HELPS TO RELEASE THE MATURE PHAGE PARTICLES FROM THE  
 CC CELL WALL BY BREAKING DOWN THE PEPTIDOGLYCAN.  
 CC -1- CATALYTIC ACTIVITY: HYDROLYSES OF THE 1,4-BETA-LINKAGES BETWEEN  
 CC N-ACETYL-D-GLUCOSAMINE AND N-ACETYL-MURAMIC ACID IN PEPTIDOGLYCAN  
 CC HETEROPOLYMERS OF THE PROKARYOTES CELL WALLS.  
 CC -1- SIMILARITY: BELONGS TO FAMILY 24 OF GLYCOSYL HYDROLASES.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL: M11813; AAA88492.1;  
 DR PIR: A26225; WMBP15.  
 DR HSP: A26225; WMBP15.  
 DR PFAM: PF00959; Phage\_lysozyme; 1.  
 DR PFAM: PF01476; P3\_binding\_2; 2.  
 DR Late Protein; Hydrolase; Glycosidase; Bacteriolytic enzyme.  
 DR ACT SITE 15 PROTON DONOR (BY SIMILARITY).  
 FT SEQUENCE 258 AA: 28052 MW: 1457FC5334637169 CRC64:  
 SQ  
 Query Match 51.9% Score 55: DB 1: Length 258:  
 Best Local Similarity 50.0% Pred. No. 2: 77-00:  
 Matches 7: Conservative 5: Mismatches 2: Indels 0: Gaps 0:  
 DB 70 KALKVSVTCNQFDA 83  
 QY 1 KSMKVTVAFNQGP 14  
 RESULT 14  
 ID MP12 AMBAR STANDARD: PRT: 398 AA.  
 AC P27750:  
 DT 01-AUG-1992 (Rel. 23. Created)  
 DT 01-AUG-1992 (Rel. 23. Last sequence update)  
 DT 15-DEC-1998 (Rel. 37. Last annotation update)  
 DE POLLEN ALLERGEN AMB A 1.2 PRECURSOR (ANTIGEN E) (ANTIGEN AMB A 1).  
 GN Ambrosia artemisiifolia (short ragweed).  
 OS Ambrosia artemisiifolia: Streptophyta; Embryophyta; Tracheophyta.  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta.  
 OC eukaryotes; Spermatophyta; Magnoliophyta; eudicotyledons;  
 OC core eudicots; Asterales; Asteraceae; Ambrosia.  
 CC Asteroideae; Heliantheae; Ambrosia.  
 CC [1]  
 CC SEQUENCE FROM N.A.  
 CC TISSUE-POLLEN;  
 CC MEDLINE: 91093235.  
 RA Rafnar T., Griffith I.J., Kuo M.-C., Bond J.F., Rogers B.L.,  
 RA Klapper D.G.:  
 RI "Cloning of Amb A 1 (antigen E), the major allergen family of short  
 RI ragweed pollen.";  
 RL J. Biol. Chem. 266:1229-1236(1991).  
 CC [2]  
 CC SEQUENCE FROM N.A., AND VARIANTS.  
 CC TISSUE-POLLEN;  
 CC MEDLINE: 92234570.  
 RA Griffith I.J., Pollock J., Klapper D.G., Rogers B.L., Nault A.K.:  
 RI "Sequence polymorphism of Amb A 1 and Amb A 11, the major allergens  
 RI in *Ambrosia artemisiifolia* (short ragweed).";  
 RL Int. Arch. Allergy Appl. Immunol. 96:296-304(1991).  
 CC -1- SUBUNIT: MONOMER.  
 CC -1- TISSUE SPECIFICITY: POLLEN AND FLOWERS.  
 CC -1- DISEASE: ONE OF THE MAJOR ALLERGENS OF THE RAGWEED POLLEN.  
 CC -1- SIMILARITY: BELONGS TO THE POLYSACCHARIDE LYASE FAMILY 1.  
 CC AMB A 1/AMB A 11/CRY J 1 SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL: M52981; AAA32566.1;  
 DR EMBL: M80559; AAA32667.1;  
 DR PIR: B39099; B39099.  
 DR PFAM: PF00544; pectinase\_1.  
 DR PRIN-S: PR00807; AMBALLENGEN  
 KW Antigen; Allergen; Signal; M; gene family: Polymorphism.  
 FT SIGNAL 1 25 ENTIAL.  
 FT CHAIN 26 398 1 LEN ALLENGEN AMB A 1.2.  
 FT MOD\_RES 226 226 91 MKED.  
 FT VARIANT 345 345 K -> K.  
 FT VARIANT 381 381 K -> K.  
 SQ SEQUENCE 398 AA: 43664 MW: 320076429474150 CRC64:  
 Query Match 51.9% Score 55: DB 1: Length 398:  
 Best Local Similarity 66.7% Pred. No. 2: 77-00:  
 Matches 8: Conservative 1: Mismatches 3: Indels 0: Gaps 0:  
 DB 257 KGLATVAFNMF 268  
 QY 1 KSMKVTVAFNQF 12  
 RESULT 15  
 ID YAMB-SCHPO STANDARD: PRT: 486 AA.  
 AC Q10063:  
 DT 01-FEB-1996 (Rel. 33. Created)  
 DT 01-FEB-1996 (Rel. 33. Last sequence update)  
 DT 01-FEB-1996 (Rel. 33. Last annotation update)  
 DE HYPOTHETICAL 53.9 KD PROTEIN C1F5.96C IN CHROMOSOME 1.  
 GN SPAC1F5.08C.  
 OS Schizosaccharomyces pombe (Pilsdon yeast).  
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetales;  
 OC Schizosaccharomycetaceae; Schizosaccharomycetes.  
 CC [1]  
 CC SEQUENCE FROM N.A.  
 CC STRAIN-972;  
 CC Genes S. J. O'Neil C. Church. M. Farnell. H.G. Kandelman M.A.  
 CC Walsh S.V.  
 CC Submitted (Nov-1995) to the EMBL/GenBank/DBJ databases  
 CC -1- SIMILARITY: SOME TO YEAST MDI.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL: Z68136; CAA92236.1;  
 KW Hypothetical protein; Transmembrane; Glycoprotein.  
 FT TRANSMEM 12 32 POTENTIAL.  
 FT TRANSMEM 465 485 POTENTIAL.  
 FT CARBOHYD 33 33 POTENTIAL.  
 FT CARBOHYD 49 49 POTENTIAL.  
 FT CARBOHYD 59 59 POTENTIAL.  
 FT CARBOHYD 82 82 POTENTIAL.  
 FT CARBOHYD 93 93 POTENTIAL.  
 FT CARBOHYD 160 160 POTENTIAL.  
 FT CARBOHYD 181 181 POTENTIAL.  
 FT CARBOHYD 185 185 POTENTIAL.  
 FT CARBOHYD 201 201 POTENTIAL.  
 FT CARBOHYD 255 255 POTENTIAL.  
 FT CARBOHYD 307 307 POTENTIAL.

FI CARBOHYD 330 330 POTENTIAL.  
FI CARBOHYD 341 341 POTENTIAL.  
FI CARBOHYD 353 353 POTENTIAL.  
FI CARBOHYD 379 379 POTENTIAL.  
FI CARBOHYD 445 445 POTENTIAL.  
SI SEQUENCE 486 AA: 53888 MW: 182047356008958 CR564:

Query Match 51.9% Score 55: DB 1: Length 486:  
Best Local Similarity 61.6% Pred. No: 2.77e+00:  
Matches 7: Conservative 3: Mismatches 1: Indels 0: Gaps 0:

BT 118 VSSAFSUFAPN 128  
II IIII  
QY 5 VIVAFNOFGN 15

Search completed: Mon Jun 19 16:09:08 2000  
Total time : 16 secs.



```

OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta:
OC ephyllophytes: Spermatophyta: Magnoliophyta: eudicotyledons:
OC core eudicots: Rosidae: eurosids II: Brassicales: Brassicaceae:
OC Arabidopsis.
[1]
RP SEQUENCE FROM N.A.
RA BEVAN M., WEDLER H., WAMBURTI R., HANZHOEF I., MEWES H.W., MAYER K.
RA SCHUELLER C.
RL Submitted (MAR-1998) to the EMBL/GenBank/CCDB databases.
[2]
RP SEQUENCE FROM N.A.
RA EU ARABIDOPSIS SEQUENCING PROJECT.
RL Submitted (APR-1998) to the EMBL/GenBank/CCDB databases.
DR EMBL: AL022440; CAA19111.1;
DR MENDEL: 29747; Arath:1088;29747.
DR PFAM: PF00544; pec_lyase; 1.
DR PRINTS: PR00807; AMBALLERGEN.
KW Lyase.
SQ SEQUENCE 394 AA: 43476 MW: 95499178 CPC12.

Query Match 82.1% Score 87 DB 10 Length 194
Best Local Similarity 91.7% Prd NO 1.46e-07
Matches 10: Conservative 1 Mismatches 3 Indels 0 Gaps 0

DB 256 MKVTVAFNHFGP 267
      1.....1111
QY 3 MKVTVAFNQFGP 14

RESULT 5
ID 065388 PRELIMINARY: FR1: 197 AA.
AC 065388:
DT 01-AUG-1998 (TREMBLrel_07, Created)
DI 01-AUG-1998 (TREMBLrel_07, Last sequence update)
DI 01-NOV-1998 (TREMBLrel_12, Last annotation update)
DE F12F1.22 PROTEIN.
GN F12F1.22
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta:
OC ephyllophytes: Spermatophyta: Magnoliophyta: eudicotyledons:
OC core eudicots: Rosidae: eurosids II: Brassicales: Brassicaceae:
OC Arabidopsis.
[1]
RP SEQUENCE FROM N.A.
RA STRAINOV G.
RA VLASTAKA A.
RA KVAN A.
RA DEKAR K.
RA SUN H.
RA DAVIS K.W.
RA EKER C.
RL Submitted (MAY-1999) to the EMBL/GenBank/CCDB databases.
DR EMBL: ACC02131; AAC17625.1;
DR MENDEL: 29024; Arath:1088;29024.
DR PFAM: PF00544; pec_lyase; 1.
DR PRINTS: PR00807; AMBALLERGEN.
SQ SEQUENCE 390 AA: 43354 MW: 80424FR4 CRC32:

Query Match 81.1% Score 86 DB 10 Length 190
Best Local Similarity 71.1% Prd NO 2.67e-07
Matches 10: Conservative 3 Mismatches 1 Indels 0 Gaps 0

DB 244 KIMRVTIAFNHFGP 257
      1.....1111
QY 1 KSMKVIVAFNQFGP 14

RESULT 6
ID 043783 PRELIMINARY: FR1: 398 AA.
AC 043783:
DT 01-NOV-1996 (TREMBLrel_01, Created)
DI 01-NOV-1996 (TREMBLrel_01, Last sequence update)
DI 01-NOV-1999 (TREMBLrel_12, Last annotation update)
DE PEPTIDE LYASE (EC 4.2.2.2) (PEPTATE TRANSPEPTIDINASE) (FRAGMENT).

```



```

GN PEL.
GS Musa acuminata (Banana).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida;
OC Zingiberales; Musaceae; Musa.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV, DWARF CAVENDISH; TISSUE=RIPE FRUIT;
RA DOMINGUEZ-PUIGJANER E.;
RI Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CV, DWARF CAVENDISH; TISSUE=RIPE FRUIT;
RA DOMINGUEZ-PUIGJANER E.;
RI Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: ELIMINATIVE CLEAVAGE OF PECTATE TO GIVE
CC OLIGOSACCHARIDES WITH 4-DEOXY-ALPHA-D-GLUC-4-ENURONOSYL GROUPS AT
CC THEIR NON-REDUCING ENDS.
DR EMBL: X92943; CAA63496.1;
DR MENDEL: 9136; Musac:1089;9136.
DR PFAM: PF00544; pec_lyase; 1.
KW Lyase.
FT SIGNAL: 1
FT CHAIN: 1
FT NON_TER
SQ SEQUENCE 396 AA: 4371 MW: 5598332A CRC32:

Query Match 77.4% Score 82: DB 10: Length 398:
Best Local Similarity 66.7% Pred. No. 2.87e-06:
Matches 10: Conservative 3: Mismatches 2: Indels 0: Gaps 0:

Db 258 KMKQVTIAFNHFG 272
QY 1 KSMKVTVAFNQFG 15

RESULT 7 PRELIMINARY: PRT: 226 AA.
ID Q23657
AC Q23657
DI 01-JAN-1998 (TRENBLrel: 05, Created)
DI 01-JAN-1998 (TRENBLrel: 05, Last sequence update)
DI 01-NOV-1999 (TRENBLrel: 12, Last annotation update)
DE PUTATIVE PECTATE LYASE (FRAGMENT).
GN A10.
CS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Rosidae; eustosids II; Brassicales; Brassicaceae;
OC Arabidopsis.
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE: 97422403.
RA KULIKAUSSAS R., MCCORMICK S.;
RI "Identification of the tobacco and Arabidopsis homologues of the
RI pollen-expressed LAT59 gene of tomato.";
RL Plant Mol. Biol. 34:809-814(1997).
DR EMBL: U83622; AAB69761.1;
DR MENDEL: 25561; Arabid:1088;25561.
DR PFAM: PF00544; pec_lyase; 1.
KW Lyase.
FT NON_TER
FT NON_TER
FT NON_TER
SQ SEQUENCE 226 AA: 25:03 MW: 5FLC8DD2 CRC32:

Query Match 73.6% Score 78: DB 10: Length 226:
Best Local Similarity 69.2% Pred. No. 2.96e-05:
Matches 9: Conservative 3: Mismatches 1: Indels 0: Gaps 0:

Db 196 KMKQVTIAFNHFG 208
QY 1 KSMKVTVAFNQFG 13

RESULT 9 PRELIMINARY: PRT: 459 AA.
ID Q24554
AC Q24554
DI 01-JAN-1998 (TRENBLrel: 05, Created)
DI 01-JAN-1998 (TRENBLrel: 05, Last sequence update)
DI 01-NOV-1999 (TRENBLrel: 12, Last annotation update)
DE PECTATE LYASE PRECURSOR.
OS Zinnia elegans.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Asterales; eustosids II; Asterales; Asteraceae;
OC Asteroideae; Heliantheae; Zinnia.
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE: 98345940.
RA DOMINGO C., ROBERTS K., STALEY N.J., CONNERTON J., MCCANN M.C.;
RI "A pectate lyase from Zinnia elegans is auxin inducible.";
RL Plant J. 13:17-28(1998).
DR EMBL: Y09341; CAA70735.1;
DR MENDEL: 26064; Zinel:1088;26064.
DR PFAM: PF00544; pec_lyase; 1.
DR PRINTS: PR00807; AMBALLERGEN.
KW Signal; Lyase.
FT SIGNAL: 1
FT CHAIN: 21
FT NON_TER
SQ SEQUENCE 401 AA: 44407 MW: C4C07131 CRC32:

Query Match 73.6% Score 78: DB 10: Length 401:
Best Local Similarity 69.2% Pred. No. 2.96e-05:
Matches 9: Conservative 3: Mismatches 1: Indels 0: Gaps 0:

Db 261 KMKQVTIAFNHFG 273
QY 1 KSMKVTVAFNQFG 13

RESULT 9 PRELIMINARY: PRT: 455 AA.
ID Q64510
AC Q64510
DI 01-AUG-1998 (TRENBLrel: 07, Created)
DI 01-AUG-1998 (TRENBLrel: 07, Last sequence update)
DI 01-NOV-1999 (TRENBLrel: 12, Last annotation update)
DE PUTATIVE PECTATE LYASE.
GN T20F6.14.
CS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Rosidae; eustosids II; Brassicales; Brassicaceae;
OC Arabidopsis.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV, COLUMBIA;
RA ROUNSLEY S.D., LIN X., KETCHUM K.A., CROSBY M.L., BRANDON R.C.,
RA SYKES S.M., KAUL S., MASON T.M., KERLAVAGE A.R., ADAMS M.D.,
RA SOMERVILLE C.R., VENTER J.C.;
RI Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AC002521; AAC05350.1;
DR MENDEL: 28505; Arabid:1088;28505.
DR PFAM: PF00544; pec_lyase; 1.
DR PRINTS: PR00807; AMBALLERGEN.
KW Lyase.
FT NON_TER
FT NON_TER
FT NON_TER
SQ SEQUENCE 455 AA: 51257 MW: EBA0B82D CRC32:

Query Match 71.7% Score 76: DB 10: Length 455:
Best Local Similarity 69.2% Pred. No. 9.30e-05:
Matches 9: Conservative 2: Mismatches 2: Indels 0: Gaps 0:

Db 309 KMKQVTIAFNHFG 321
QY 1 KSMKVTVAFNQFG 13

RESULT 10 PRELIMINARY: PRT: 459 AA.
ID Q23665
AC Q23665

```

01-JAN-1998 (TRENBLREL, 05, Created)  
 01-JAN-1998 (TRENBLREL, 05, Last sequence update)  
 01-NOV-1999 (TRENBLREL, 12, Last annotation update)  
 59. POTATIVE PECTATE LYASE.  
 Arabidopsis thaliana (Mouse-ear cress).  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;  
 core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;  
 Arabidopsis.  
 [1]  
 SEQUENCE FROM N.A.  
 MEDLINE: 97422433.  
 KULIKAUSSAS R., MCCORMICK S.L.  
 Identification of the tobacco and Arabidopsis homologues of the  
 poliovirus expressed LAT59 gene of tomato.  
 Plant Mol. Biol. 34:809-814 (1997).  
 EMBL: J83619; AAB69759.1; -  
 MENDEL: 25559; Arabid1088:25559.  
 PFAM: PF00544; pec\_lyase.1.  
 PRINTS: PR00807; AMBALLERGEN.  
 KW Lyase.  
 SQ SEQUENCE 459 AA: 51420 MW: 41E859D8 CRC32:  
 Query Match: 71.74; Score 75; DB 10; Length 459;  
 Best Local Similarity 69.24; Pred. No. 9; Gaps 0;  
 Matches 9; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
 01 313 KMKQITVAVNHFQ 325  
 1 1111111111  
 2 KSMKVIVAFNGFG 13  
 RESULT 11 PRELIMINARY: PRT: 431 AA.  
 AC Q23017.  
 01-JAN-1998 (TRENBLREL, 05, Created)  
 01-JAN-1998 (TRENBLREL, 05, Last sequence update)  
 01-NOV-1999 (TRENBLREL, 12, Last annotation update)  
 113117 PROTEIN.  
 Arabidopsis thaliana (Mouse-ear cress).  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;  
 core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;  
 Arabidopsis.  
 [1]  
 SEQUENCE FROM N.A.  
 STRAIN-CV, COLUMBIA;  
 OSBORNE R.L., VISHNUSKAYA V.S., TREHMI M., YU G., JI L., LIU S.,  
 LI J., HOANG L., EUEHLER E., CONWAY A.B., CONWAY A.R., DEWAR K.,  
 FENG J., KIM C., KURTZ D., LI Y., SHINN P., SUN H., DAVIS R.W.,  
 ECKER J.R., FEDERSPIEL N.A., THEOLOGIS A.  
 Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.  
 [2]  
 SEQUENCE FROM N.A.  
 STRAIN-CV, COLUMBIA;  
 THEOLOGIS A.;  
 Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.  
 [3]  
 SEQUENCE FROM N.A.  
 STRAIN-CV, COLUMBIA;  
 THEOLOGIS A.;  
 Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.  
 [4]  
 SEQUENCE FROM N.A.  
 STRAIN-CV, COLUMBIA;  
 THEOLOGIS A.;  
 Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.  
 EMBL: AC002376; AAB80622.1; -  
 MENDEL: 25798; Arabid1088:25798.  
 PFAM: PF00544; pec\_lyase.1.  
 PRINTS: PR00807; AMBALLERGEN.

SQ SEQUENCE 431 AA: 47770 MW: 86E8C285 CRC32:  
 Query Match: 70.88; Score 75; DB 10; Length 431;  
 Best Local Similarity 61.58; Pred. No. 1; Gaps 0;  
 Matches 8; Conservative 4; Mismatches 1; Indels 0; Gaps 0;  
 01 289 KMQVTIAYNHFG 301  
 1111111111  
 2 KSMKVIVAFNGFG 13  
 RESULT 12 PRELIMINARY: PRT: 253 AA.  
 AC Q4084.  
 01-NOV-1996 (TRENBLREL, 01, Created)  
 01-NOV-1996 (TRENBLREL, 01, Last sequence update)  
 01-NOV-1999 (TRENBLREL, 12, Last annotation update)  
 PECTATE LYASE (EC 4.2.2.2) (FRAGMENT).  
 Nicotiana tabacum (Common tobacco).  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;  
 core eudicots; Asteridae; euasterids II; Solanales; Solanaceae;  
 Nicotiana.  
 [1]  
 SEQUENCE FROM N.A.  
 STRAIN-SAMSUN;  
 ROGERS H.J., LONSDALE D.M.;  
 Submitted (JUL-1991) to the EMBL/GenBank/DBJ databases.  
 CC -1- CATALYTIC ACTIVITY: ELIMINATIVE CLEAVAGE OF PECTATE TO GIVE  
 OLIGOSACCHARIDES WITH 4-DEOXY-ALPHA-D-GULO-4-ENOL-5,6-GLUCO-5-AL  
 CC -2- THEIR NON-REDUCING ENDS.  
 CC -3- DEVELOPMENTAL STAGE: EXPRESSED DURING MICROSPOROGENESIS.  
 CC -4- SIMILARITY: TO TOMATO PROTEIN P55 PRECURSOR (AC P15721).  
 EMBL: X61101; CAA34413.1;  
 DR MENDEL: 16497; Nicot1088:1649  
 DR PFAM: PF00544; pec\_lyase.1.  
 KW Lyase.  
 FT NON-TER  
 SQ SEQUENCE 263 AA: 29449 MW: 44826A CRC32:  
 Query Match: 69.88; Score 74; DB 10; Length 263;  
 Best Local Similarity 61.58; Pred. No. 2; Gaps 0;  
 Matches 8; Conservative 4; Mismatches 1; Indels 0; Gaps 0;  
 01 136 KMKITVAYNHFG 148  
 1111111111  
 2 KSMKVIVAFNGFG 14  
 RESULT 13 PRELIMINARY: PRT: 450 AA.  
 AC Q40319.  
 01-NOV-1996 (TRENBLREL, 01, Created)  
 01-NOV-1996 (TRENBLREL, 01, Last sequence update)  
 01-NOV-1999 (TRENBLREL, 12, Last annotation update)  
 PECTATE LYASE HOMOLOG.  
 Medicago sativa (Alfalfa).  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;  
 core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae;  
 Medicago.  
 [1]  
 SEQUENCE FROM N.A.  
 WU-Y., QIU X., DU S., ERICKSON L.;  
 Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.  
 EMBL: U41472; AAA86241.1; -  
 MENDEL: 9042; Medsa:1088:9042.  
 DR PFAM: PF00544; pec\_lyase.1.  
 DR PRINTS: PR00807; AMBALLERGEN.  
 KW Lyase.  
 SQ SEQUENCE 450 AA: 50155 MW: B35FCF9E CRC32:  
 Query Match: 69.88; Score 74; DB 10; Length 450;

Best Local Similarity 69.2% Pred. NO. 2 88e-04;  
Matches 9; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

DB 309 KIMQVTVAFNHFQ 321  
1 1111111111  
QY 1 KSMKVTVAFNQFG 13

RESULT 14  
ID 043862 PRELIMINARY: PRT: 435 AA.  
AC 043862:  
C1 01-NOV-1996 (TRENBLrel. 01; Created)  
C2 01-NOV-1996 (TRENBLrel. 01; Last sequence update)  
C3 01-NOV-1999 (TRENBLrel. 12; Last annotation update)  
DE POLLEN SPECIFIC PECTATE LYASE HOMOLOGUE (EC 4.2.2.2)  
DS (PECTATE TRANSGLUCINASE).  
OS Zea mays (Maize).  
CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
CC euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;  
CC Poaceae; Zea.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CV\_GOLDUP; TISSUE=SHOOT;  
RX MEDLINE: 94083558  
RA TURCICH M.P., HAMILTON D.A., MASCARENHAS J.P.;  
R1 "Isolation and characterization of pollen-specific maize genes with  
R1 sequence homology to ragweed allergens and pectate lyases.";  
R1 Plant Mol. Biol. 23:1061-1065(1993).  
CC -1- CATALYTIC ACTIVITY: ELIMINATIVE CLEAVAGE OF PECTATE TO GIVE  
CC OLIGOSACCHARIDES WITH 4-DEOXY-ALPHA-D-GLUC-4-ENURONOSYL GROUPS AT  
CC THEIR NON-REDUCING ENDS.  
DR EMBL: L20140; AAA16476.1;  
DR MENDEL: 11165; Zeam1:1098;11165.  
LR PFAM: PF00544; pec\_lyase; 1.  
DR PRINTS: PRCC007; AMBALLERGEN.  
KW Lyase.  
SQ SEQUENCE 438 AA: 49138 MW: 10565369 C:132;

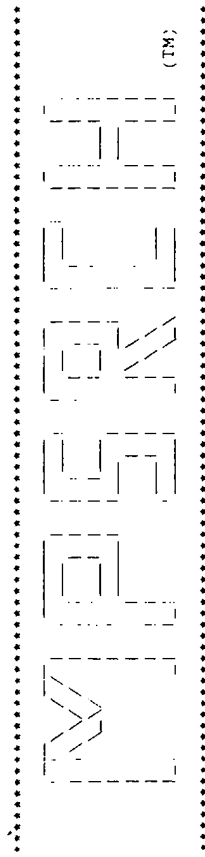
Query Match 67.9% Score 72; DB 10; Length 439;  
Best Local Similarity 91.8% Pred. NO. 6 82e-04;  
Matches 9; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

DB 309 KIMQVTVAFNHFQ 310  
1 1111111111  
QY 3 KIMQVTVAFNHFQ 13

RESULT 15  
ID 024416 PRELIMINARY: PRT: 405 AA.  
AC 024416:  
C1 01-JAN-1998 (TRENBLrel. 05; Created)  
C2 01-JAN-1998 (TRENBLrel. 05; Last sequence update)  
C3 01-NOV-1999 (TRENBLrel. 12; Last annotation update)  
DE PECTATE LYASE.  
OS P.  
CC Fragaria ananassa (Strawberry).  
CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
CC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;  
CC core eudicots; Rosidae; eurosids 1; Rosales; Rosaceae; Fragaria.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CHANDLER;  
RX MEDLINE: 97435972  
RA MEDINA-ESCOBAR N., CARDENAS J., MOYANO E., CABALLERO J.L.,  
RA MUNOZ-BLANCO J.;  
RT "Cloning, molecular characterization and expression pattern of a  
RT strawberry ripening-specific cDNA with sequence homology to pectate  
RT lyase from higher plants.";  
R1 Plant Mol. Biol. 34:867-877(1997).  
DR EMBL: J63550; AAB71208.1;  
DR MENDEL: 26059; Frag1:1088;26059.  
LR PFAM: PF00544; pec\_lyase; 1.  
KW Lyase.

SQ SEQUENCE 405 AA: 45744 MW: 289544A0 CRC32;  
Query Match 67.0% Score 71; DB 10; Length 405;  
Best Local Similarity 61.5% Pred. NO. 1 53e-03;  
Matches 8; Conservative Mismatches 2; Indels 0; Gaps 0;  
DB 309 KIMQVTVAFNHFQ 321  
1 1111111111  
QY 1 KSMKVTVAFNQFG 13

Search completed: Mon Jun 19 16:03:37 2000  
Job time : 12 secs.



Release 3.1A John F. Collins, BioComputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MFSEARCH protein - protein database search, using Smith-Waterman algorithm

Run on: Mon Jun 19 16:12:29 2000: MasPar time 3.59 Seconds  
210.856 Million cell updates/sec

Tabular output not generated.

Title: >US-09-142-524A-10  
Description: (132) from US09142524A.pep  
Perfect Score: 427  
Sequence: 1 EYLLSARDVLAWSKRMVIVAFNFGPN 32

Scoring table: PAM 150  
Gap 11

Searched: 189963 seqs, 23686106 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: a-geneseq35  
l-geneseqp

Statistics: Mean 23.654: Variance 92.032 scale 0.257

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

| SUMMARY    |       |                       |     | SUMMARY                 |           |            |       |
|------------|-------|-----------------------|-----|-------------------------|-----------|------------|-------|
| Result No. | Score | Query Match Length DB | ID  | Description             | Pred. No. | Result No. | Score |
| 1          | 114   | 50.2                  | 93  | M. leprae GroES-like p  | 2.57e-01  | 1          | 114   |
| 2          | 103   | 45.4                  | 102 | GroES structural prote  | 5.04e-02  | 2          | 103   |
| 3          | 98    | 43.2                  | 15  | Japanese cedar pollen   | 2.57e-01  | 3          | 98    |
| 4          | 98    | 43.2                  | 15  | Residues 211-225 of Cr  | 2.57e-01  | 4          | 98    |
| 5          | 98    | 43.2                  | 17  | Sugi allergen protein   | 2.57e-01  | 5          | 98    |
| 6          | 98    | 43.2                  | 20  | T-cell epitope peptide  | 2.57e-01  | 6          | 98    |
| 7          | 98    | 43.2                  | 20  | Cry j I Japanese Cedar  | 2.57e-01  | 7          | 98    |
| 8          | 98    | 43.2                  | 20  | Cry j I pollen allergen | 2.57e-01  | 8          | 98    |
| 9          | 98    | 43.2                  | 26  | Cry j I pollen allergen | 2.57e-01  | 9          | 98    |
| 10         | 98    | 43.2                  | 26  | Cry j I pollen allergen | 2.57e-01  | 10         | 98    |
| 11         | 98    | 43.2                  | 26  | Cry j I pollen allergen | 2.57e-01  | 11         | 98    |
| 12         | 98    | 43.2                  | 28  | Cry j I pollen allergen | 2.57e-01  | 12         | 98    |
| 13         | 98    | 43.2                  | 28  | Cry j I pollen allergen | 2.57e-01  | 13         | 98    |
| 14         | 98    | 43.2                  | 28  | Cry j I pollen allergen | 2.57e-01  | 14         | 98    |
| 15         | 98    | 43.2                  | 30  | Cry j I pollen allergen | 2.57e-01  | 15         | 98    |
| 16         | 98    | 43.2                  | 30  | T-cell epitope peptide  | 2.57e-01  | 16         | 98    |
| 17         | 98    | 43.2                  | 30  | Cry j I pollen allergen | 2.57e-01  | 17         | 98    |
| 18         | 98    | 43.2                  | 36  | Cry j I pollen allergen | 2.57e-01  | 18         | 98    |
| 19         | 98    | 43.2                  | 353 | Japanese cedar pollen   | 2.57e-01  | 19         | 98    |
| 20         | 98    | 43.2                  | 353 | Cedar pollen allergen   | 2.57e-01  | 20         | 98    |
| 21         | 98    | 43.2                  | 354 | Japanese cypress polle  | 2.57e-01  | 21         | 98    |
| 22         | 98    | 43.2                  | 354 | Chamaecyparis obtusa p  | 2.57e-01  | 22         | 98    |
| 23         | 98    | 43.2                  | 367 | Jun S I.                | 2.57e-01  | 23         | 98    |

PS 03-SEP-1991: F00701.  
 PR 10-SEP-1990: FR-011186.  
 PA (INSP ) INST. PASTEUR.  
 PI Mazodier P, Guichet M, G.  
 DR WPI: 92-114/58/14.  
 CP NSDB: 022483.  
 PF Recombinant DNA contd. heat inducible promoter and heterologous  
 PI gene - also vectors, transformed cells and new heat shock  
 PI proteins of Streptococcus albus  
 PS Disclosure: Fig 5: 50pp: French.  
 CC The sequence is that of the GROES protein which is encoded by the  
 CC structural gene groES. See also Q22477-022486.  
 SQ Sequence 102 AA:

Query Match 45.4% Score 103: DB 1: Length 102:  
 Best Local Similarity 76.5% Pred. No. 9,04e-02:  
 Matches 13: Conservative 3: Mismatches 1: Indels 0: Gaps 0:

DB 86 EYVLVSARDVLAIIEK 102  
 QY 1 EYVLVSARDVLAVVSK 17  
 : |||||:|||||:

RESULT 3  
 ID R89293 standard: peptide: 15 AA.  
 AC R89293:  
 DT 12-MAR-1996 (first entry)  
 DE Japanese cedar pollen allergen Cry I derived T-cell epitope peptide.  
 RE Japanese cedar: pollen allergen: Cry I: T-cell epitope: peptides:  
 KW prevention: treatment: cryptomeria pollenosis.  
 CS Cryptomeria japonica.  
 PN J07118295-A.  
 PD 09-MAY-1995.  
 PF 20-OCT-1993: 262626.  
 PR 20-OCT-1993: JP-262626.  
 RA (MEIP ) MEIJ: MILK PROD CO LTD.  
 DR WPI: 95-203834/27.  
 PI New cryptomeria pollen allergen T-cell epitope peptide - used for  
 PI prevention, treatment and investigation of Japanese cedar pollenosis  
 PS Claim 5: Page 2: 8pp: Japanese.  
 CC R75388 is the Japanese cedar pollen allergen Cry I, from which the  
 CC T-cell epitope peptides R89289-R89295 were derived. The peptides  
 CC can be used for the prevention and treatment of cryptomeria  
 CC pollenosis, and also for the investigation of pollenosis.  
 SQ Sequence 15 AA:

Query Match 43.2% Score 98: DB 1: Length 15:  
 Best Local Similarity 86.7% Pred. No. 2,57e-01:  
 Matches 13: Conservative 1: Mismatches 1: Indels 0: Gaps 0:

DB 1 KSKKVTVAFNQFGPN 15  
 QY 18 RRMKVTVAFNQFGPN 32  
 : |||||:|||||:

RESULT 4  
 ID W57755 standard: peptide: 15 AA.  
 AC W57755:  
 DT 17-SEP-1996 (first entry)  
 DE Residues 211-225 of Cry j 1.  
 KW Cry j 1: Japanese cedar pollen antigen: allergy: immunotherapy:  
 CS HLA class II molecule.  
 CS Cryptomeria japonica.  
 PN W09820902-A1.  
 PD 22-MAY-1998.  
 PF 12-NOV-1997: J04129.  
 PR 13-NOV-1996: JP-302053.  
 RA (MEIP ) MEIJ: MILK PROD CO LTD.  
 PI Dairiki K, Kikuo K, Kume A, Sone T:  
 DR WPI: 98-29767/26.  
 PI Peptides derived from Japanese cedar pollen antigens are  
 PI immunotherapeutic agents - useful for allergy treatment and typing  
 PI HLA class II molecules in allergy sufferers

PS Claim 12: Page 28: 50pp: Japanese.  
 CC This sequence represents residues 211-225 of the Cry j 1 protein, and  
 CC is a peptide of the invention. The peptides are derived from Japanese  
 CC cedar pollen antigens, and are used as immunotherapeutic agents in the  
 CC treatment of allergy. The peptides can be used for identification and  
 CC typing of the particular HLA class II molecules in an allergy sufferer,  
 CC and also for peptide immunotherapy of an allergy. Using these peptides  
 CC the immunotherapy can be targeted more specifically to the requirements  
 CC of the individual patient, allowing more effective treatment of an  
 CC allergy, including those patients for whom treatment with a conventional  
 CC immunotherapeutic agent is ineffective.  
 SQ Sequence 15 AA:

Query Match 43.2% Score 98: DB 1: Length 15:  
 Best Local Similarity 86.7% Pred. No. 2,57e-01:  
 Matches 13: Conservative 1: Mismatches 1: Indels 0: Gaps 0:

DB 1 KSKKVTVAFNQFGPN 15  
 QY 18 RRMKVTVAFNQFGPN 32  
 : |||||:|||||:

RESULT 5  
 ID W80349 standard: peptide: 17 AA.  
 AC W80349:  
 DT 11-JAN-1999 (first entry)  
 DE Sugi allergen protein Cryj2 derived epitope for T cells.  
 KW T cell epitope: sugi allergen proteins Cryj1: Cryj2: treatment:  
 KW sugi-pollenosis: allergic reaction: pollen.  
 CS Synthetic.  
 PN J0259198-A.  
 PD 29-SEP-1998.  
 PF 22-DEC-1997: 353448.  
 PR 24-DEC-1996: JP-34341.  
 RA (HAYB ) HAYASHIBARA SEIBUTSU KAGAKO.  
 RA (SANY ) SANKYO CO LTD.  
 DR WPI: 98-577037/49.  
 PI A linked T cell epitope peptide - used for the treatment of  
 PI sugi-pollenosis  
 PS Claim 7: Page 18: 21pp: Japanese.  
 CC W80339-58 represent epitopes for T cells, derived from the sugi allergen  
 CC proteins Cryj1 (W80339-44, W80350-53 and W80356-58) and Cryj2 (W80345-49  
 CC and W80354-55). The peptides are useful for the treatment of  
 CC sugi-pollenosis, an allergic reaction of the body to pollen.  
 SQ Sequence 17 AA:

Query Match 43.2% Score 98: DB 1: Length 17:  
 Best Local Similarity 85.7% Pred. No. 2,57e-01:  
 Matches 13: Conservative 1: Mismatches 1: Indels 0: Gaps 0:

DB 2 KSKKVTVAFNQFGPN 16  
 QY 18 RRMKVTVAFNQFGPN 32  
 : |||||:|||||:

RESULT 6  
 ID W42144 standard: peptide: 20 AA.  
 AC W42144:  
 DT 16-JUN-1998 (first entry)  
 DE T-cell epitope peptide 24 from Japanese cypress pollen antigen: Chao1.  
 KW Japanese cypress pollen: antigen: T-cell epitope: Chao1: Chao2:  
 KW diagnosis: allergy: spring tree pollen disease: pollinosis.  
 CS Chamaecyparis obtusa.  
 PN W09747648-A1.  
 PD 18-DEC-1997.  
 PF 12-JUN-1997: J02031.  
 PR 14-JUN-1996: JP-153527.  
 RA (MEIP ) MEIJ: MILK PROD CO LTD.  
 PI Dairiki K, Kikuo K:  
 DR WPI: 98-052242/05.  
 PI T-cell epitope peptide portion of Japanese cypress pollen antigens  
 PI Chao1 and Chao2 - used for diagnosis and treatment of spring tree  
 PI pollen disease

PS Claim 1: Page 27-28: 71pp: Japanese.  
 CC The present sequence represents a T-cell epitope peptide from Japanese  
 CC cypress pollen antigen Chaoi. The present invention describes peptides  
 CC which correspond to the T-cell epitope sites on Japanese cypress pollen  
 CC antigens Chaoi and Chao2. The peptides can be used as a reagent for the  
 CC diagnosis of allergy to Japanese cypress pollen, and as an antigen in  
 CC the treatment and prevention of spring tree pollen disease in which the  
 CC pollinosis involves reactivity to Japanese cypress pollen.  
 CC Sequence 20 AA:

Query Match 43.2% Score 98: DB 1: Length 20:  
 Best Local Similarity 86.7%: Pred. No. 2,578-01:  
 Matches 13: Conservative 1: Mismatches 1: Indels 0: Gaps 0:

Db 1 KSMKVTVAFNFGPN 15  
 : ||| ||||| ||  
 QY 18 RRMKVTVAFNFGPN 32

RESULT 7  
 ID R82512 standard: Protein: 20 AA.  
 AC R82512:  
 DT 15-APR-1996 (first entry)  
 DE Cry j 1 Japanese cedar pollen allergen peptide fragment (CJI-22).  
 KW Cry j 1: Japanese cedar pollen allergen; modified: drug production;  
 KW allergy: Cryptomeria japonica.  
 OS Cryptomeria japonica.  
 PN W09527786-A1.  
 PD 19-OCT-1995.  
 PF 06-APR-1995: U04245.  
 PR 08-APR-1994: US-226248.  
 PR 06-DEC-1994: US-350225.  
 PA (IMMU-) IMMUNOLOGIC PHARM CORP.  
 PI Chen X, Evans S, Franzen HM, Kuo M, Powers SP:  
 PI Skated 2;  
 DR WPI: 95-366391/47.  
 PT Modified Cryptomeria japonica (Cry j) 1 peptide(s) - useful for  
 PT treating allergy to Japanese cedar pollen allergen or  
 PT immunologically cross reactive allergens  
 PS Disclosure: Figure 2: 60pp: English.  
 CC Novel peptides of cry j 1 have been modified as a part of a  
 CC therapeutic scheme to develop an optimised drug product for  
 CC cedar pollen allergen of an allergen which is immunologically cross  
 CC reactive with Japanese cedar pollen allergen. Such modified peptides  
 CC possess certain characteristics which render them particularly  
 CC suitable for drug product formulation. Peptide fragments of Cry j 1,  
 CC modified and unmodified, are given in 82491-R82525. This peptide  
 CC fragment corresponds to amino acids 211-230 of the allergen mature  
 CC protein.  
 CC Sequence 20 AA:

Query Match 43.2% Score 98: DB 1: Length 20:  
 Best Local Similarity 86.7%: Pred. No. 2,578-01:  
 Matches 13: Conservative 1: Mismatches 1: Indels 0: Gaps 0:

Db 1 KSMKVTVAFNFGPN 15  
 : ||| ||||| ||  
 QY 18 RRMKVTVAFNFGPN 32

RESULT 8  
 ID R45563 standard: Protein: 20 AA.  
 AC R45563:  
 DT 13-JUL-1994 (first entry)  
 DE Cry j 1 pollen allergen peptide CJI-22.  
 KW Japanese cedar: detection; allergy; treatment; diagnosis;  
 KW T cell epitope; sensitivity.  
 OS Cryptomeria japonica.  
 PN W09401540-A.  
 PD 20-JAN-1994.  
 PF 15-JAN-1993: U00139.  
 PR 10-JUL-1992: WO-U05561.

PR 01-SEP-1992: US-938990.  
 PA (IMMU-) IMMUNOLOGIC PHARM CORP.  
 PI Bond JF, Garman RD, Griffith J3, Kuo M, Pollock J:  
 DR WPI: 94-035066/04.  
 DE Antigen derived from Japanese cedar pollen allergen Cry j 1  
 DE contain at least two T cell epitope(s), used to treat or diagnose  
 DE allergy  
 PS Claim 1: Fig 13: 137pp: English.  
 CC The sequence is that of an isolated peptide of the Japanese cedar  
 CC pollen allergen Cry j 1 (amino acids 211-230). The peptide, CJI-22,  
 CC can be used for the treatment and diagnosis of allergies associated  
 CC with Japanese cedar pollen. It has enhanced therapeutic properties  
 CC but reduced side effects compared to naturally occurring allergens  
 CC Sequence 20 AA:

Query Match 43.2% Score 98: DB 1: Length 20:  
 Best Local Similarity 86.7%: Pred. No. 2,578-01:  
 Matches 13: Conservative 1: Mismatches 1: Indels 0: Gaps 0:

Db 1 KSMKVTVAFNFGPN 15  
 : ||| ||||| ||  
 QY 18 RRMKVTVAFNFGPN 32

RESULT 9  
 ID R45592 standard: Protein: 26 AA.  
 AC R45592:  
 DT 13-JUL-1994 (first entry)  
 DE Cry j 1 pollen allergen peptide CJI-43.10.  
 KW Japanese cedar: detection; allergy; treatment; diagnosis;  
 KW T cell epitope; sensitivity.  
 OS Cryptomeria japonica.  
 PN W09401560-A.  
 PD 20-JAN-1994.  
 PF 15-JAN-1993: U00139.  
 PR 10-JUL-1992: WO-U05561.  
 PR 01-SEP-1992: US-938990.  
 PA (IMMU-) IMMUNOLOGIC PHARM CORP.  
 PI Bond JF, Garman RD, Griffith J3, Kuo M, Pollock J:  
 DR WPI: 94-03 066/04.  
 DE Antigen derived from Japanese cedar pollen allergen Cry j 1  
 DE contain at least two T cell epitope(s), used to treat or diagnose  
 DE allergy  
 PS Claim 76: Fig 18: 137pp: English.  
 CC The sequence is that of an isolated peptide of the Japanese cedar  
 CC pollen allergen Cry j 1. The peptide, CJI-43.10, can be used for  
 CC the treatment and diagnosis of allergies associated with Japanese  
 CC cedar pollen. It has enhanced therapeutic properties but reduced  
 CC side effects compared to naturally occurring allergens.  
 CC Sequence 26 AA:

Query Match 43.2% Score 98: DB 1: Length 26:  
 Best Local Similarity 86.7%: Pred. No. 2,578-01:  
 Matches 13: Conservative 1: Mismatches 1: Indels 0: Gaps 0:

Db 1 KSMKVTVAFNFGPN 15  
 : ||| ||||| ||  
 QY 18 RRMKVTVAFNFGPN 32

RESULT 10  
 ID R45594 standard: Protein: 26 AA.  
 AC R45594:  
 DT 13-JUL-1994 (first entry)  
 DE Cry j 1 pollen allergen peptide CJI-43.12.  
 KW Japanese cedar: detection; allergy; treatment; diagnosis;  
 KW T cell epitope; sensitivity.  
 OS Cryptomeria japonica.  
 PN W09401560-A.  
 PD 20-JAN-1994.  
 PF 15-JAN-1993: U00139.  
 PR 10-JUL-1992: WO-U05561.  
 PR 01-SEP-1992: US-938990.

PA (IMMU-) IMMUNOLOGIC PHARM CORP.  
 PI Bond JF, Garman RD, Griffith LJ, Kuo M, Pollock J;  
 CR WPI: 94-035066/04.  
 DE Antigen derived from Japanese cedar pollen allergen Cry j 1.  
 KW Contain at least two T cell epitope(s), used to treat or diagnose  
 PT allergy.  
 PS Claim 76: Fig 18: 137pp: English.  
 CC The sequence is that of an isolated peptide of the Japanese cedar  
 CC pollen allergen Cry j 1. The peptide, CJI-43.8, can be used for  
 CC the treatment and diagnosis of allergies associated with Japanese  
 CC cedar pollen. It has enhanced therapeutic properties but reduced  
 CC side effects compared to naturally occurring allergens.  
 SQ Sequence 26 AA.

Query Match 43.2% Score 98 DB 1 Length 26;  
 Best Local Similarity 86.7% Pred. No. 2,576-01;  
 Matches 13: Conservative 1: Mismatches 1: Indels 0: Gaps 0;

DB 1 KSMKVTVAFNQCPN 15  
 : ||||| |||||  
 QY 18 RRMKVTVAFNQCPN 32

RESULT 11  
 ID R45591 standard: Protein: 26 AA.  
 AC R45591;  
 DT 13-JUL-1994 (first entry)  
 DE Cry j 1 pollen allergen peptide CJI-43.8.  
 KW Japanese cedar; detection: allergy; treatment: diagnosis;  
 KW T cell epitope; sensitivity.  
 OS Cryptomeria japonica.  
 PN W09401560-A.  
 PD 20-JAN-1994.  
 PF 15-JAN-1993: U00139.  
 PR 10-JUL-1992: WO-005661.  
 PS 01-SEP-1992: US-938990.  
 PA (IMMU-) IMMUNOLOGIC PHARM CORP.  
 PI Bond JF, Garman RD, Griffith LJ, Kuo M, Pollock J;  
 CR WPI: 94-035066/04.  
 DE Antigen derived from Japanese cedar pollen allergen Cry j 1.  
 KW Contain at least two T cell epitope(s), used to treat or diagnose  
 PT allergy.  
 PS Claim 76: Fig 18: 137pp: English.  
 CC The sequence is that of an isolated peptide of the Japanese cedar  
 CC pollen allergen Cry j 1. The peptide, CJI-43.8, can be used for  
 CC the treatment and diagnosis of allergies associated with Japanese  
 CC cedar pollen. It has enhanced therapeutic properties but reduced  
 CC side effects compared to naturally occurring allergens.  
 SQ Sequence 26 AA.

Query Match 43.2% Score 98 DB 1 Length 26;  
 Best Local Similarity 86.7% Pred. No. 2,576-01;  
 Matches 13: Conservative 1: Mismatches 1: Indels 0: Gaps 0;

DB 1 KSMKVTVAFNQCPN 15  
 : ||||| |||||  
 QY 18 RRMKVTVAFNQCPN 32

RESULT 12  
 ID R45590 standard: Protein: 26 AA.  
 AC R45590;  
 DT 13-JUL-1994 (first entry)  
 DE Cry j 1 pollen allergen peptide CJI-43.8.  
 KW Japanese cedar; detection: allergy; treatment: diagnosis;  
 KW T cell epitope; sensitivity.  
 OS Cryptomeria japonica.  
 PN W09401560-A.  
 PD 20-JAN-1994.  
 PF 15-JAN-1993: U00139.  
 PR 10-JUL-1992: WO-005661.  
 PS 01-SEP-1992: US-938990.  
 PA (IMMU-) IMMUNOLOGIC PHARM CORP.

PI Bond JF, Garman RD, Griffith LJ, Kuo M, Pollock J;  
 CR WPI: 94-035066/04.  
 DE Antigen derived from Japanese cedar pollen allergen Cry j 1.  
 KW Contain at least two T cell epitope(s), used to treat or diagnose  
 PT allergy.  
 PS Claim 76: Fig 18: 137pp: English.  
 CC The sequence is that of an isolated peptide of the Japanese cedar  
 CC pollen allergen Cry j 1. The peptide, CJI-43.8, can be used for  
 CC the treatment and diagnosis of allergies associated with Japanese  
 CC cedar pollen. It has enhanced therapeutic properties but reduced  
 CC side effects compared to naturally occurring allergens.  
 SQ Sequence 28 AA.

Query Match 43.2% Score 98 DB 1 Length 28;  
 Best Local Similarity 86.7% Pred. No. 2,576-01;  
 Matches 13: Conservative 1: Mismatches 1: Indels 0: Gaps 0;

DB 1 KSMKVTVAFNQCPN 15  
 : ||||| |||||  
 QY 18 RRMKVTVAFNQCPN 32

RESULT 13  
 ID R45589 standard: Protein: 28 AA.  
 AC R45589;  
 DT 13-JUL-1994 (first entry)  
 DE Cry j 1 pollen allergen peptide CJI-43.7.  
 KW Japanese cedar; detection: allergy; treatment: diagnosis;  
 KW T cell epitope; sensitivity.  
 OS Cryptomeria japonica.  
 PN W09401560-A.  
 PD 20-JAN-1994.  
 PF 15-JAN-1993: U00139.  
 PR 10-JUL-1992: WO-005661.  
 PS 01-SEP-1992: US-938990.  
 PA (IMMU-) IMMUNOLOGIC PHARM CORP.  
 PI Bond JF, Garman RD, Griffith LJ, Kuo M, Pollock J;  
 CR WPI: 94-035066/04.  
 DE Antigen derived from Japanese cedar pollen allergen Cry j 1.  
 KW Contain at least two T cell epitope(s), used to treat or diagnose  
 PT allergy.  
 PS Claim 76: Fig 18: 137pp: English.  
 CC The sequence is that of an isolated peptide of the Japanese cedar  
 CC pollen allergen Cry j 1. The peptide, CJI-43.7, can be used for  
 CC the treatment and diagnosis of allergies associated with Japanese  
 CC cedar pollen. It has enhanced therapeutic properties but reduced  
 CC side effects compared to naturally occurring allergens.  
 SQ Sequence 28 AA.

Query Match 43.2% Score 98 DB 1 Length 28;  
 Best Local Similarity 86.7% Pred. No. 2,576-01;  
 Matches 13: Conservative 1: Mismatches 1: Indels 0: Gaps 0;

DB 1 KSMKVTVAFNQCPN 15  
 : ||||| |||||  
 QY 18 RRMKVTVAFNQCPN 32

RESULT 14  
 ID R45593 standard: Protein: 28 AA.  
 AC R45593;  
 DT 13-JUL-1994 (first entry)  
 DE Cry j 1 pollen allergen peptide CJI-43.11.  
 KW Japanese cedar; detection: allergy; treatment: diagnosis;  
 KW T cell epitope; sensitivity.  
 OS Cryptomeria japonica.  
 PN W09401560-A.  
 PD 20-JAN-1994.  
 PF 15-JAN-1993: U00139.  
 PR 10-JUL-1992: WO-005661.  
 PS 01-SEP-1992: US-938990.  
 PA (IMMU-) IMMUNOLOGIC PHARM CORP.  
 PI Bond JF, Garman RD, Griffith LJ, Kuo M, Pollock J;

DR WPI: 94-035086/34.  
PI Antigens derived from Japanese cedar pollen allergen Cry j 1 -  
PI contain at least two T cell epitope(s), used to treat or diagnose  
PI allergy.  
PS Claim 76: Fig 18: 137pp: English.  
CC The sequence is that of an isolated peptide of the Japanese cedar  
CC pollen allergen Cry j 1. The peptide, CJI-43.11, can be used for  
CC the treatment and diagnosis of allergies associated with Japanese  
CC cedar pollen. It has enhanced therapeutic properties but reduced  
CC side effects compared to naturally occurring allergens.  
SQ Sequence 28 AA:

Query Match 43.2%, Score 98: DB 1: Length 30:  
Best Local Similarity 86.7%: Pred No. 2,578-01:  
Matches 13: Conservative 1: Mismatches 1: Indels 0: Gaps 0:

EB 1 KSMKVTVAFNFGPN 15  
QY 18 PRMKVTVAFNFGPN 32

RESULT 15  
ID R45588 standard: Protein: 10 AA.  
AC R45588:  
DE 13-JUL-1994 (first entry)  
DE Cry j 1 pollen allergen peptide CJI-43.6.  
KW Japanese cedar: detection; allergy: treatment; diagnosis;  
KW T cell epitope: sensitivity.  
QS Cryptomeria japonica  
PN M06401560-A.  
PD 20-JAN-1994:  
PF 15-JAN-1993: U00139.  
PR 10-JUL-1992: M0-005661.  
PR 01-SEP-1992: US-93890.  
PA (IMMU-) IMMUNOLOGIC PHARM CORP.  
PI Bond JF, Garman RD, Griffith TJ, Kuo M, Pollock J:  
DR WPI: 94-035086/34.  
PI Antigens derived from Japanese cedar pollen allergen Cry j 1 -  
PI contain at least two T cell epitope(s), used to treat or diagnose  
PI allergy.  
PS Claim 76: Fig 18: 137pp: English.  
CC The sequence is that of an isolated peptide of the Japanese cedar  
CC pollen allergen Cry j 1. The peptide, CJI-43.6, can be used for  
CC the treatment and diagnosis of allergies associated with Japanese  
CC cedar pollen. It has enhanced therapeutic properties but reduced  
CC side effects compared to naturally occurring allergens.  
SQ Sequence 30 AA:

Query Match 43.2%, Score 98: DB 1: Length 30:  
Best Local Similarity 86.7%: Pred No. 2,578-01:  
Matches 13: Conservative 1: Mismatches 1: Indels 0: Gaps 0:

EB 1 KSMKVTVAFNFGPN 15  
QY 18 PRMKVTVAFNFGPN 32

Search completed: Mon Jun 19 16:12:39 2000  
Job time : 11 secs.



\*\*\*\*\*

W P E E H (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Tue Jun 20 13:36:17 2000: MasPar time 2.72 Seconds  
tabular output not generated. 169,949 Million cell updates/sec

Title: >US-09-142-524A-10  
Description: (1-32) from US09142524A.pep  
Perfect Score: 227  
Sequence: 1: EEYILSARDVLAAYSKRMKVTVAFNQFGPN 32

Scoring table: PAM 150  
Gap 11

Searched: 145341 seqs, 14437480 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: a-issued  
1:5A\_COMB 2:5B\_COMB 3:6\_COMB 4:PCI\_COMB 5:backfiles:

Statistics: Mean 22.169; Variance 87.947; scale 0.252

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score the result being printed,  
and is derived by analysis of the total scores distribution.

SUMMARIES

| Result No. | Query | Score | Match | Length | ID         | Description             | Pred. No. |
|------------|-------|-------|-------|--------|------------|-------------------------|-----------|
| 1          | 114   | 50.2  | 93    | 2      | US-08-467- | Sequence 36, Applicatio | 4,56e-01  |
| 2          | 72    | 31.7  | 2050  | 2      | US-08-347- | Sequence 2, Applicatio  | 2,43e-01  |
| 3          | 71    | 31.3  | 558   | 4      | PCI-US93-  | Sequence 2, Applicatio  | 2,95e-01  |
| 4          | 70    | 30.8  | 94    | 2      | US-08-467- | Sequence 38, Applicatio | 3,57e-01  |
| 5          | 70    | 30.8  | 1895  | 2      | US-08-619- | Sequence 2, Applicatio  | 3,57e-01  |
| 6          | 70    | 30.8  | 2813  | 3      | US-08-896- | Sequence 2, Applicatio  | 3,57e-01  |
| 7          | 68    | 30.0  | 254   | 2      | US-09-144- | Sequence 23, Applicati  | 5,21e-01  |
| 8          | 68    | 30.0  | 254   | 2      | US-08-685- | Sequence 23, Applicati  | 5,21e-01  |
| 9          | 68    | 30.0  | 269   | 2      | US-07-857- | Sequence 79, Applicati  | 5,21e-01  |
| 10         | 68    | 30.0  | 304   | 2      | US-08-701- | Sequence 27, Applicati  | 5,21e-01  |
| 11         | 68    | 30.0  | 316   | 2      | US-08-838- | Sequence 15, Applicati  | 5,21e-01  |
| 12         | 68    | 30.0  | 316   | 1      | US-08-278- | Sequence 16, Applicati  | 5,21e-01  |
| 13         | 67    | 29.5  | 368   | 2      | US-08-824- | Sequence 3, Applicatio  | 6,30e-01  |
| 14         | 67    | 29.5  | 737   | 1      | US-08-119- | Sequence 5, Applicatio  | 6,30e-01  |
| 15         | 67    | 29.5  | 737   | 3      | US-08-335- | Sequence 4, Applicatio  | 6,30e-01  |
| 16         | 67    | 29.5  | 1687  | 2      | US-08-570- | Sequence 29, Applicati  | 6,30e-01  |
| 17         | 67    | 29.5  | 1704  | 3      | US-08-336- | Sequence 10, Applicati  | 6,30e-01  |
| 18         | 67    | 29.5  | 3387  | 2      | US-08-473- | Sequence 7, Applicatio  | 6,30e-01  |
| 19         | 66    | 29.1  | 1876  | 2      | US-08-619- | Sequence 2, Applicatio  | 7,60e-01  |
| 20         | 65    | 28.6  | 94    | 2      | US-08-467- | Sequence 39, Applicati  | 9,16e-01  |
| 21         | 65    | 28.6  | 97    | 2      | US-08-467- | Sequence 40, Applicati  | 9,16e-01  |
| 22         | 65    | 28.6  | 988   | 2      | US-08-286- | Sequence 15, Applicati  | 9,16e-01  |
| 23         | 65    | 28.6  | 988   | 3      | US-08-980- | Sequence 15, Applicati  | 9,16e-01  |

|    |    |      |      |   |            |                        |          |
|----|----|------|------|---|------------|------------------------|----------|
| 24 | 65 | 28.6 | 1052 | 2 | US-08-852- | Sequence 2, Applicatio | 9,16e-01 |
| 25 | 64 | 28.2 | 291  | 2 | US-08-838- | Sequence 6, Applicatio | 1,10e-02 |
| 26 | 64 | 28.2 | 359  | 1 | US-07-562- | Sequence 4, Applicatio | 1,10e-02 |
| 27 | 64 | 28.2 | 501  | 4 | PCI-US95-  | Sequence 9, Applicatio | 1,10e-02 |
| 28 | 64 | 28.2 | 501  | 2 | US-08-023- | Sequence 9, Applicatio | 1,10e-02 |
| 29 | 64 | 28.2 | 501  | 2 | US-07-553- | Sequence 9, Applicatio | 1,10e-02 |
| 30 | 64 | 28.2 | 501  | 2 | US-08-362- | Sequence 9, Applicatio | 1,10e-02 |
| 31 | 64 | 28.2 | 501  | 2 | US-08-288- | Sequence 9, Applicatio | 1,10e-02 |
| 32 | 64 | 28.2 | 616  | 2 | US-01-001- | Sequence 5, Applicatio | 1,10e-02 |
| 33 | 64 | 28.2 | 686  | 2 | US-03-993- | Sequence 12, Applicati | 1,10e-02 |
| 34 | 63 | 27.8 | 540  | 1 | US-08-216- | Sequence 31, Applicati | 1,10e-02 |
| 35 | 63 | 27.8 | 797  | 4 | PCI-US95-  | Sequence 2, Applicatio | 1,10e-02 |
| 36 | 63 | 27.8 | 797  | 2 | US-08-352- | Sequence 2, Applicatio | 1,10e-02 |
| 37 | 63 | 27.8 | 973  | 1 | US-08-162- | Sequence 8, Applicatio | 1,10e-02 |
| 38 | 63 | 27.8 | 1012 | 1 | US-08-219- | Sequence 1, Applicatio | 1,10e-02 |
| 39 | 63 | 27.8 | 1012 | 1 | US-08-219- | Sequence 8, Applicatio | 1,10e-02 |
| 40 | 63 | 27.8 | 1012 | 1 | US-08-219- | Sequence 9, Applicatio | 1,10e-02 |
| 41 | 63 | 27.8 | 1012 | 1 | US-08-219- | Sequence 5, Applicatio | 1,10e-02 |
| 42 | 63 | 27.8 | 1012 | 1 | US-07-944- | Sequence 2, Applicatio | 1,10e-02 |
| 43 | 63 | 27.8 | 1012 | 1 | US-08-219- | Sequence 4, Applicatio | 1,10e-02 |
| 44 | 63 | 27.8 | 1012 | 1 | US-08-219- | Sequence 2, Applicatio | 1,10e-02 |
| 45 | 63 | 27.8 | 1012 | 1 | US-08-219- | Sequence 14, Applicati | 1,10e-02 |

ALIGNMENTS

RESULT 1  
ID US-08-467-822-36 STANDARD: PRT: 93 AA.  
XX  
AC  
XX  
XX  
DI

Sequence 36, Application US/08457622

Sequence 36, Application US/08457622  
Patent No. 5843450

GENERAL INFORMATION:

APPLICANT: Labigne, Agnes  
APPLICANT: Sauerbaum, Sebastian  
APPLICANT: Ferrero, Richard L.  
APPLICANT: Tiberge, Jean-Michel  
TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS AGAINST  
TITLE OF INVENTION: HELICOBACTER INFECTION, POLYPEPTIDES FOR USE IN THE  
TITLE OF INVENTION: COMPOSITIONS, AND NUCLEIC ACID SEQUENTS INCLUDING SAID  
TITLE OF INVENTION: POLYPEPTIDES  
NUMBER OF SEQUENCES: 46  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Labigne, Agnes, 1400 1<sup>st</sup> Street, N.W.  
STREET: 1400 1<sup>st</sup> Street, N.W.  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20005-3315  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/467,822  
FILING DATE: 05-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/447,177  
FILING DATE: 19-MAY-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/432,697  
FILING DATE: 02-MAY-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:

NAME: Meyers, Kenneth J.  
 REGISTRATION NUMBER: 25,146  
 REFERENCE/DOCKET NUMBER: 03495.0137-2000  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (202) 408-4000  
 TELEFAX: (202) 408-4400  
 INFORMATION FOR SEQ ID NO: 36:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 93 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 SEQUENCE 93 AA: 10002 MW: 48030 CN:  
 EVYLILSARCVLAIVSK 93  
 EVYLILSARCVLAIVSK 17

Query Match 50.2% Score 114: DB 2: Length 93:  
 Best Local Similarity 100.0% Pred. No. 4,56e-01:  
 Matches 17: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

DE 77 EVYLILSARCVLAIVSK 93  
 QY 1 EVYLILSARCVLAIVSK 17

RESULT 2  
 ID US-09-347-594A-2 STANDARD: PRT: 2050 AA:  
 XX xxxxxx  
 DT  
 DE  
 DE Sequence 2, Application US/08347594A

Sequence 2, Application US/08347594A  
 Patent No. 5849536  
 GENERAL INFORMATION:  
 APPLICANT: Gartinkel, Leonard  
 APPLICANT: Richter, Zafar  
 TITLE OF INVENTION: CLONING AND PRODUCTION OF HUMAN VON  
 TITLE OF INVENTION: WALLBRAND FACTOR GP12 BINDING DOMAIN POLYPEPTIDES AND  
 TITLE OF INVENTION: METHODS OF USING SAME  
 NUMBER OF SEQUENCES: 4  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: John P. White  
 STREET: 1185 Avenue of the Americas  
 CITY: New York  
 STATE: New York  
 COUNTRY: USA  
 ZIP: 10046  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: floppy disk  
 COMPUTER: IBM PC COMPATIBLE  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: WORDPERFECT 5.1  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: PCT/US93/0027  
 FILING DATE: 19930401  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US/57/864,987  
 FILING DATE: 05-APR-1992  
 ATTORNEY/AGENT INFORMATION:  
 NAME: DOROTHY R. ADH  
 REGISTRATION NUMBER: P-35,414  
 REFERENCE/DOCKET NUMBER: 2050-101,171  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 212-278-0400  
 TELEFAX: 212-278-0400  
 INFORMATION FOR SEQ ID NO: 2:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 2050 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 SEQUENCE 2050 AA: 225692 MW: 21477294 CN:

Query Match 31.7% Score 72: DB 2: Length 2050:  
 Best Local Similarity 34.6% Pred. No. 2,43e-01:  
 Matches 9: Conservative 6: Mismatches 9: Indels 0: Gaps 0:

DB 198 YLILGKALSVVWDRHLSISVVKQ 223  
 QY 3 YLILSARCVLAIVSKRKVTVAFNQ 28

RESULT 3  
 ID PCT-US93-03027-2 STANDARD: PRT: 558 AA:  
 XX xxxxxx  
 DT  
 DE  
 DE Sequence 2, Application PCT/US9303027

Sequence 2, Application PCT/US9303027  
 GENERAL INFORMATION:  
 APPLICANT: LEONARD, WARREN: ILLERMAN,  
 APPLICANT: MICHEL  
 TITLE OF INVENTION: CONTROL AND/OR  
 TITLE OF INVENTION: PREVENTION OF BINDING OF NF- $\kappa$ B/REL-1: KSAI  
 TITLE OF INVENTION:  
 NUMBER OF SEQUENCES: 9  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: MORGAN & FINNEGAN  
 STREET: 345 PARK AVENUE  
 CITY: NEW YORK  
 STATE: NEW YORK  
 COUNTRY: USA  
 ZIP: 10154  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: FLOPPY DISK  
 COMPUTER: IBM PC COMPATIBLE  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: WORDPERFECT 5.1  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: PCT/US93/0027  
 FILING DATE: 19930401  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US/57/864,987  
 FILING DATE: 05-APR-1992  
 ATTORNEY/AGENT INFORMATION:  
 NAME: DOROTHY R. ADH  
 REGISTRATION NUMBER: P-35,414  
 REFERENCE/DOCKET NUMBER: 2050-101,171  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 212-278-0400  
 TELEFAX: 212-278-0400  
 INFORMATION FOR SEQ ID NO: 1:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 558  
 TYPE: AMINO ACID  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 HYPOTHETICAL: NO  
 ORIGINAL SOURCE:  
 ORGANISM: mouse  
 STRAIN:  
 INDIVIDUAL ISOLATE:  
 DEVELOPMENTAL STAGE:  
 HAPLOTYPE:  
 TISSUE TYPE:  
 CELL TYPE: fibroblast  
 CELL LINE: NIH-3T3  
 ORGANISLE:  
 FEATURE:  
 NAME/KEY: Relb  
 LOCATION:

```

CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 08/432,697
CC FILING DATE: 02-MAY-1995
CC CLASSIFICATION: 435
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Meyers, Kenneth J.
CC REGISTRATION NUMBER: 25,146
CC REFERENCE/DOCKET NUMBER: G3495.0137-02000
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (202) 408-4000
CC TELEFAX: (202) 408-4400
CC INFORMATION FOR SEQ ID NO. 38:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 94 amino acids
CC TYPE: amino acid
CC STR/NDEONESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC SEQUENCE 94 AA: 10343 MW: 45288 CN:
CC
CC Query Match 30.8% Score 70 DP 2 Length 94
CC Best Local Similarity 57.1% Pred No. 3 57e-G1:
CC Matches 8 Conservative 3 Mismatches 3 Indels 0
CC
Cc Db 80 EYILRESDI:LAVI 93
Cc Qy 2 EYILSARDV:LAVV 15
Cc
Cc RESULT 5
Cc ID US-08-619-554-4 STANDARD: PRI: 1895 AA.
Cc AC xxxxxx
Cc XX
Cc D:
Cc XX
Cc XX
Cc DE Sequence 4, Application US/08619554
Cc DX
Cc CC Sequence 4, Application US/08619554
Cc CC Patent No. 5821353
Cc CC GENERAL INVENTION:
Cc CC APPLICANT: DOUGLAS, Carl Jon M.
Cc CC APPLICANT: CHREBET, Gary L.
Cc CC APPLICANT: CLEMAS, Joseph
Cc CC APPLICANT: EL-SHERBEINI, Mohammed
Cc CC APPLICANT: FOOR, Forrest
Cc CC APPLICANT: KAHN, Jennifer.
Cc CC APPLICANT: KELLY, Rosengate, - PARENT, S.A.
Cc CC APPLICANT: MARRINAN, Jason - RAHMAN, N.M.
Cc CC APPLICANT: MORIN, Nancy - REISTEPEL, E.A.
Cc CC APPLICANT: ONISHI, Junzo - SHEN, Guo Yu
Cc CC TITLE OF INVENTION: DNA ENCODING 1,3 BETA-D-GALACTAN
Cc CC TITLE OF INVENTION: SYNTHASE SUBUNIT5
Cc CC NUMBER OF SEQUENCES: 8
Cc CC CORRESPONDENCE ADDRESS:
Cc CC ADDRESSEE: JOSEPH A. COPPOLA - MERCK & CO., INC.
Cc CC STREET: 126 EAST LINCOLN AVENUE - P.O. BOX 2000
Cc CC CITY: RAHWAY
Cc CC STATE: NJ
Cc CC COUNTRY: USA
Cc CC ZIP: 07065
Cc CC COMPUTER READABLE FORM:
Cc CC MEDIUM TYPE: Diskette
Cc CC COMPUTER: IBM Compatib.
Cc CC OPERATING SYSTEM: DOS
Cc CC SOFTWARE: FastSeq for Windows Version 2.0
Cc CC CURRENT APPLICATION DATA:
Cc CC APPLICATION NUMBER: US/08/619,554
Cc CC FILING DATE: 01-AUG-1995
Cc CC CLASSIFICATION: 536
Cc CC PRIOR APPLICATION DATA:
Cc CC APPLICATION NUMBER:
Cc CC FILING DATE:

```

CC ATTORNEY/AGENT INFORMATION:  
CC NAME: COPPOLA, JOSEPH A.  
CC REGISTRATION NUMBER: 38,413  
CC REFERENCE/DOCKET NUMBER: 19104PI  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: 732-594-6734  
CC TELEFAX: 732-594-4720  
CC TELEX:  
CC INFORMATION FOR SEQ ID NO: 4:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 1895 amino acids  
CC TYPE: amino acid  
CC STRANDEDNESS: unknown  
CC TOPOLOGY: unknown  
CC MOLECULE TYPE: protein  
CC SEQUENCE 1895 AA: 216588 MW: 18982148 CN:  
SQ  
Query Match 30.8% Score 70: DB 2: Length 1895:  
Best Local Similarity 47.5% Pred. NO. 3.57e-01:  
Matches 9: Conservative 6: Mismatches 9: Indels 0: Gaps 0:  
DB 658 FLLSLKGFIRKLTSTSMRCITGKY 681  
QY 3 EYLISARDVLAVVSKRPMKVTVAF 26  
RESULT 6  
ID US-08-896-449A-2 STANDARD: PRI: 2813 AA:  
XX  
AC xxxxxx  
DE  
Sequence 2, Application US/08896449A  
Patent No. 640143  
GENERAL INFORMATION:  
APPLICANT: Vectra, Patrick J.  
APPLICANT: Yuzbysyan-Gurkhan, Vilma  
APPLICANT: Schall, William D.  
APPLICANT: Brewer, George J.  
TITLE OF INVENTION: DNA ENCODING CANINE VON WILLEBRAND  
TITLE OF INVENTION: FACTOR AND METHODS OF USE  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Haines, Dickey & Pierce, P.C.  
STREET: 5445 Corporate Drive  
CITY: Troy  
STATE: Michigan  
COUNTRY: USA  
ZIP: 48068  
COMPUTER READABLE FORM  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOCS/MS-DOCS  
SOFTWARE: PatentIt Release 1.0, Version #1.25  
CURRENT APPLICATION DATA:  
FILING DATE: 18-JUL-1997  
APPLICATION NUMBER: US/08/896.449A  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Smith, DeAnn F.  
REFERENCE/DOCKET NUMBER: 2115-001226  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 248-641-1600  
TELEFAX: 248-641-0270  
TELEX: 387637  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2813 amino acids  
TYPE: amino acid  
TOPOLOGY: linear

CC MOLECULE TYPE: protein  
CC SEQUENCE 2813 AA: 309523 : 40158915 CN:  
SQ  
Query Match 30.8% Score 70: DB 3: Length 2813:  
Best Local Similarity 29.8% Pred. NO. 3.57e-01:  
Matches 8: Conservative 11: Mismatches 9: Indels 0: Gaps 0:  
DB 960 QYVILLGNALSVVMDHRLSISVTLKR 986  
QY 2 EYLISARDVLAVVSKRPMKVTVAF 28  
RESULT 7  
ID US-09-144-925-23 STANDARD: PRI: 254 AA:  
XX  
AC xxxxxx  
DE  
Sequence 23, Application US/09144925  
Patent No. 5951979  
GENERAL INFORMATION:  
APPLICANT: Toks, Nicholas  
APPLICANT: Flint, Andrew J.  
TITLE OF INVENTION: SUBSTRATE TRAPPING PROTEIN  
TITLE OF INVENTION: TYROSINE PHOSPHATASES  
NUMBER OF SEQUENCES: 36  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: HAMILTON, B. K. SMITH & REYNOLDS, P.C.  
STREET: Two Militia Dr.  
CITY: Lexington  
STATE: MA  
COUNTRY: USA  
ZIP: 02421-4799  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: Windows 95  
SOFTWARE: FastSeq for Windows Version 2.00  
CURRENT APPLICATION DATA:  
FILING DATE: 05/09/1995  
APPLICATION NUMBER: US/09/144.925  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/85.992  
FILING DATE: July 25, 1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Sheridan, Patricia  
REGISTRATION NUMBER: 3227  
REFERENCE/DOCKET NUMBER: 058-41-0  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 781-861-0240  
TELEFAX: 781-861-9540  
TELEX:  
INFORMATION FOR SEQ ID NO: 23:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 254 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
SEQUENCE 254 AA: 26034 MW: 364632 CN:  
SQ  
Query Match 30.0% Score 68: DB 2: Length 254:  
Best Local Similarity 46.7% Pred. NO. 5.21e-01:  
Matches 7: Conservative 4: Mismatches 4: Indels 0: Gaps 0:  
DB 230 IQRKMKMVTQTTFQ 244  
QY 15 VSKRPMKVTVAFNQF 29

[illegible][illegible]

```

CC CC NUMBER OF SEQUENCES: 41
CC CC CORRESPONDENCE ADDRESS:
CC CC ADDRESSEE: Lyon & Lyon
CC CC STREET: 613 West Fifth Street
CC CC CITY: Suite 4700
CC CC STATE: Los Angeles
CC CC COUNTRY: California
CC CC ZIP: 90013-2066
CC CC COMPUTER READABLE FORM:
CC CC MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
CC CC NAME: Kurdydk, Linda M.
CC CC REGISTRATION NUMBER: 34,971
CC CC OPERATING SYSTEM: IBM PC, DOS 5.0
CC CC SOFTWARE: FASTSEQ for Windows 2.0
CC CC CURRENT APPLICATION DATA:
CC CC APPLICATION NUMBER: US/08/701.191A
CC CC FILING DATE: August 21, 1996
CC CC CLASSIFICATION: 530
CC CC PRIOR APPLICATION DATA:
CC CC APPLICATION NUMBER:
CC CC FILING DATE:
CC CC ATTORNEY/AGENT INFORMATION:
CC CC NAME: Warburg, Richard J.
CC CC REGISTRATION NUMBER: 32,327
CC CC REFERENCE/DOCKET NUMBER: 227/488
CC CC TELECOMMUNICATION INFORMATION:
CC CC TELEPHONE: (213) 489-1600
CC CC TELEFAX: (213) 955-0440
CC CC TELEX: 67-3510
CC CC INFORMATION FOR SEQ ID NO: 27:
CC CC SEQUENCE CHARACTERISTICS:
CC CC LENGTH: 304 amino acids
CC CC TYPE: amino acid
CC CC STRANDEDNESS: single
CC CC TOPOLOGY: linear
CC CC MOLECULE TYPE: protein
CC CC IMMEDIATE SOURCE:
CC CC CLONE: Ret
CC CC SEQUENCE 304 AA: 34556 MW: 480234 CN:
Query Match 30.0% Score 68: DB 2: Length 304:
Best Local Similarity 44.4% Pred. No. 5,21: -01:
Matches 8: Conservative 6: Mismatches 3: Indels 1: Gaps 1:
DB 165 LAARNIL-VAEGRKMKIS 161
QY 6 LSARDVLAVVSKRMRKVT 23
RESULT 11
ID US-08-348-057A-15 STANDARD 165 316 AA
AC xxxxxx
XX
XX
XX
XX
Sequence 15: Application US/086338957A
DE Patent No. 5998157
CC GENERAL INFORMATION:
CC APPLICANT: Breitman, Martin L.
CC APPLICANT: Rossant, Janet
CC APPLICANT: Dumont, Daniel J.
CC APPLICANT: Yamaguchi, Terry P.
CC TITLE OF INVENTION: NO. 5998187e1 Receptor Tyrosine Kinase
CC NUMBER OF SEQUENCES: 32
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Bereskin & Parr
CC STREET: 40 King Street West
CC CITY: Toronto
CC STATE: Ontario
CC COUNTRY: Canada
CC ZIP: M5H 3Y2
CC
CC Sequence 16: Application US/08278089A
DE Patent No. 56817-4
CC GENERAL INFORMATION:
CC APPLICANT: Breitman, Martin L.
CC APPLICANT: Rossant, Janet
CC APPLICANT: Yamaguchi, Terry P.
CC TITLE OF INVENTION: NO. 56817-4 Receptor Tyrosine Kinase
CC NUMBER OF SEQUENCES: 34
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Bereskin & Parr
CC STREET: 40 King Street West
CC CITY: Toronto
CC STATE: Ontario
CC COUNTRY: Canada
CC ZIP: M5H 3Y2
CC
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patent's Release #1.0, Version #1.30
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/838.957A
CC FILING DATE: 23-APR-1997
CC CLASSIFICATION: 435
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Kurdydk, Linda M.
CC REGISTRATION NUMBER: 34,971
CC REFERENCE/DOCKET NUMBER: 3153-212
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (416) 364-7311
CC TELEFAX: (416) 361-1398
CC INFORMATION FOR SEQ ID NO: 15:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 316 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC IMMEDIATE SOURCE:
CC CLONE: Ret
CC SEQUENCE 316 AA: 35696 MW: 1-0817 CN:
Query Match 30.0% Score 68: DB 2: Length 316:
Best Local Similarity 44.4% Pred. No. 5,21e-01:
Matches 8: Conservative 6: Mismatches 3: Indels 1: Gaps 1:
DB 152 LAARNIL-VAEGRKMKIS 168
QY 6 LSARDVLAVVSKRMRKVT 23
RESULT 12
ID US-08-278-089A-16 STANDARD 167 316 AA
AC xxxxxx
XX
XX
XX
XX
Sequence 16: Application US/08278089A
DE Patent No. 56817-4
CC GENERAL INFORMATION:
CC APPLICANT: Breitman, Martin L.
CC APPLICANT: Rossant, Janet
CC APPLICANT: Yamaguchi, Terry P.
CC TITLE OF INVENTION: NO. 56817-4 Receptor Tyrosine Kinase
CC NUMBER OF SEQUENCES: 34
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Bereskin & Parr
CC STREET: 40 King Street West
CC CITY: Toronto
CC STATE: Ontario
CC COUNTRY: Canada
CC ZIP: M5H 3Y2
CC
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patent's Release #1.0, Version #1.30
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/278.089A
CC FILING DATE: 20-JUL-1994
CC CLASSIFICATION: 530
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Kurdydk, Linda M.
CC REGISTRATION NUMBER: 34,971
CC

```

|  |        |   |   |              |
|--|--------|---|---|--------------|
|  | CC     | LIBRARY:  | GeoBank   |              |
|  | CC     | CLONE:  | 205276  |              |
|  | SQ     | SEQUENCE  | 368 AA; 39086 MW; 716242 CN;                        |              |
|  |        | Query Match   | 29.5%; Score 67; DB 2; Length 368;                  |              |
|  |        | Best Local Similarity   | 32.3%; Pred. No. 6,30e-01;                          |              |
|  |        | Matches   | 10: Conservative ; Mismatches 10: Indels 1: Gaps 1: |              |
|  | Dd     | 283   | ELEKTVKNCLADDERKLK-SIAFSPIS 312                     |              |
|  | Oy     | 1   | EYLILSARDVLAVSKRMKVAFNFGP 31                        |              |
|  |        |   | :        :  |              |
|  | RESULT | 14  |   |              |
|  | ID     | US-08-119-361-5   | STANDARD:   | PMT: 737 AA. |
|  | XX     | xxxxxx  |   |              |
|  | DI     | XX  |   |              |
|  | DE     | Sequence 5, Application US/08119361                                       |   |              |
|  | XX     | Sequence 5, Application US/08119361                                       |   |              |
|  | CC     | Patent No. 5523390  |   |              |
|  | CC     | GENERAL INFORMATION:  |   |              |
|  | CC     | APPLICANT: Travis, James  |   |              |
|  | CC     | APPLICANT: Potempa, Jan   |   |              |
|  | CC     | APPLICANT: Barr, Philip J.  |   |              |
|  | CC     | TITLE OF INVENTION: Porphyromonas gingivalis Arginine-specific Proteinase |   |              |
|  | CC     | NUMBER OF SEQUENCES: 16   |   |              |
|  | CC     | CORRESPONDENCE ADDRESS:   |   |              |
|  | CC     | ADDRESSEE: Greenlee and Winner, P.C.                                      |   |              |
|  | CC     | STREET: 5370 Manhattan Circle, Suite 201                                  |   |              |
|  | CC     | CITY: Boulder   |   |              |
|  | CC     | STATE: CO   |   |              |
|  | CC     | COUNTRY: USA  |   |              |
|  | CC     | ZIP: 80303  |   |              |
|  | CC     | COMPUTER READABLE FORM:   |   |              |
|  | CC     | MEDIUM TYPE: Floppy diskette  |   |              |
|  | CC     | COMPUTER: IBM PC compatible   |   |              |
|  | CC     | OPERATING SYSTEM: PC-DOS/MS-DOS   |   |              |
|  | CC     | SOFTWARE: Patent In Release #1.0, Version #1.25                           |   |              |
|  | CC     | CURRENT APPLICATION DATA:   |   |              |
|  | CC     | APPLICATION NUMBER: US/08/119,361   |   |              |
|  | CC     | FILING DATE: 10-SEP-1993  |   |              |
|  | CC     | CLASSIFICATION: A24   |   |              |
|  | CC     | ATTORNEY/AGENT INFORMATION:   |   |              |
|  | CC     | NAMES: Ferber, Dorcas E.  |   |              |
|  | CC     | REGISTRATION NUMBER: 44784  |   |              |
|  | CC     | REFERENCE/DOCKET NUMBER: 21144  |   |              |
|  | CC     | TELECOMMUNICATION INFORMATION:  |   |              |
|  | CC     | TELEPHONE: 303-459-8080   |   |              |
|  | CC     | TELEFAX: 303-499-8089   |   |              |
|  | CC     | INFORMATION FOR SEQ ID NO: 5:   |   |              |
|  | CC     | SEQUENCE CHARACTERISTICS:   |   |              |
|  | CC     | LENGTH: 737 amino acids   |   |              |
|  | CC     | TYPE: amino acid  |   |              |
|  | CC     | TOPOLOGY: linear  |   |              |
|  | CC     | MOLECULE TYPE: protein  |   |              |
|  | SQ     | SEQUENCE 737 AA: 81332 MW; 1874143 CN;                                    |   |              |
|  |        | Query Match   | 29.5%; Score 67; DB 1; Length: 737;                 |              |
|  |        | Best Local Similarity   | 52.0%; Pred. No. 6,30e-01;                          |              |
|  |        | Matches   | 13: Conservative ; Mismatches 9: Indels 2: Gaps 2:  |              |
|  | Dd     | 87  | IILS-RS-LAVSDIREMKVEVYSKF 109                       |              |
|  | Oy     | 5   | IILSARDVLAVSKRMKVTFVNFI 29                          |              |
|  |        |   | :   |              |
|  | RESULT | 15  |   |              |
|  | ID     | US-08-336-308A-4  | STANDARD:   | PPT: 737 AA  |

XX  
AC  
XX  
DT  
XX  
XX

xxxxxx

Sequence 4, Application US/08336308A

Sequence 4, Application US/08336308A  
Patent No. 6017532

GENERAL INFORMATION:

APPLICANT: Travis, James

APPLICANT: Potempa, Jan S.

APPLICANT: Barr, Philip J.

APPLICANT: Pavloff, Nadine

TITLE OF INVENTION: Forphyromonas gingivalis

TITLE OF INVENTION: Arginine-specific Proteinase Coding Sequences

NUMBER OF SEQUENCES: 16

CORRESPONDENCE ADDRESS:

ADDRESSEE: Greenlee, Wimer and Sullivan, P.C.

STREET: 500 Manhattan Circle, Suite 201

CITY: Boulder

STATE: Colorado

COUNTRY: US

ZIP: 80303

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08336308A

FILING DATE: 08-NOV-1994

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/119,361

FILING DATE: 10-SEP-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/265,441

FILING DATE: 24-JUN-1994

ATTORNEY/AGENT INFORMATION:

NAME: Feiler, Donna M.

REGISTRATION NUMBER: 33,878

REFERENCE/DOCKET NUMBER: 21-930

TELECOMMUNICATION INFORMATION:

TELEPHONE: (303) 499-8080

TELEFAX: (303) 499-8089

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 737 amino acids

TYPE: amino acid

LOCALITY: linear

MOLECULE TYPE: protein

SEQUENCE 737 AA: 8132 MW, 2874143 CN:

Query Match 29.5% score 57; DB 3; Length 737;

Best Local Similarity 52.0%; Pred. No. 6,30e-01;

Matches 13; Conservative 1; Mismatches 9; Indels 2; Gaps 2;

Db 87 ILS-RS-LAVSNTREMKVSVSKF 109

5 ILSARDV-LAVSKRMKVIVAFNQF 29

Search completed: Tue Jun 20 13:36:23 2000

Job time : 6 secs.



\*\*\*\*\*  
W P E E A  
\*\*\*\*\*  
(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit,  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPSrch\_PP protein - protein database search, using Smith-Waterman algorithm  
Run on: Mon Jun 19 16:26:34 2000; MasPar time 17.22 Seconds  
187.279 Million cell updates/sec  
Tabular output not generated.

Title: >US-09-142-524A-10  
Description: (1-32) from US09142524A.pep  
Sequence: 1 EBYLILSARDV-AVWSKRMKVIVAFNQFNP 32

Scoring table: PAM 150  
Gap 11

Searched: 721208 seqs, 100765575 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: 3-pending  
1:PC1 2:U6 3:U60 4:U7 5:U80 6:U81 7:U92 8:U83 9:U84A  
10:U84B 11:U85 12:U86 13:U87 14:U88 15:U89 16:U90 17:U91  
18:U92 19:U93 20:U94 21:U95 22:NEWP 23:NEWC60 24:NEWC8  
25:NEWC9

Statistics: Mean 26.635; Variance 90.120; Scale 0.296

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description                       | Pred. No. |
|------------|-------|-------------|--------|----|-----------------------------------|-----------|
| 1          | 114   | 50.2        | 93     | 10 | US-08-466- Sequence 36, Applicat  | 1.68e-02  |
| 2          | 114   | 50.2        | 93     | 9  | US-08-432- Sequence 36, Applicat  | 1.68e-02  |
| 3          | 105   | 46.3        | 13     | 10 | US-08-468- Sequence 236, Applicat | 1.39e-01  |
| 4          | 105   | 46.3        | 19     | 8  | US-08-350- Sequence 236, Applicat | 1.39e-01  |
| 5          | 105   | 46.3        | 19     | 10 | US-08-467- Sequence 236, Applicat | 1.39e-01  |
| 6          | 105   | 46.3        | 19     | 10 | US-08-467- Sequence 236, Applicat | 1.39e-01  |
| 7          | 105   | 46.3        | 19     | 10 | US-08-467- Sequence 236, Applicat | 1.39e-01  |
| 8          | 105   | 46.3        | 29     | 8  | US-08-458- Sequence 222, Applicat | 1.39e-01  |
| 9          | 105   | 46.3        | 29     | 8  | US-08-458- Sequence 222, Applicat | 1.39e-01  |
| 10         | 105   | 46.3        | 29     | 10 | US-08-467- Sequence 222, Applicat | 1.39e-01  |
| 11         | 105   | 46.3        | 29     | 10 | US-08-467- Sequence 222, Applicat | 1.39e-01  |
| 12         | 105   | 46.3        | 29     | 10 | US-08-467- Sequence 222, Applicat | 1.39e-01  |
| 13         | 103   | 45.4        | 26     | 8  | US-08-350- Sequence 228, Applicat | 2.21e-01  |
| 14         | 103   | 45.4        | 26     | 8  | US-08-350- Sequence 229, Applicat | 2.21e-01  |
| 15         | 103   | 45.4        | 26     | 10 | US-08-467- Sequence 229, Applicat | 2.21e-01  |
| 16         | 103   | 45.4        | 26     | 10 | US-08-467- Sequence 228, Applicat | 2.21e-01  |
| 17         | 103   | 45.4        | 26     | 10 | US-08-468- Sequence 228, Applicat | 2.21e-01  |
| 18         | 103   | 45.4        | 26     | 10 | US-08-468- Sequence 229, Applicat | 2.21e-01  |
| 19         | 103   | 45.4        | 26     | 10 | US-08-467- Sequence 228, Applicat | 2.21e-01  |
| 20         | 103   | 45.4        | 26     | 10 | US-08-467- Sequence 229, Applicat | 2.21e-01  |

|    |     |      |     |    |                                   |          |
|----|-----|------|-----|----|-----------------------------------|----------|
| 21 | 103 | 45.4 | 26  | 10 | US-08-467- Sequence 228, Applicat | 2.21e-01 |
| 22 | 103 | 45.4 | 26  | 10 | 'S-08-467- Sequence 229, Applicat | 2.21e-01 |
| 23 | 98  | 43.2 | 19  | 10 | 'S-08-467- Sequence 229, Applicat | 6.99e-01 |
| 24 | 98  | 43.2 | 20  | 7  | US-08-226- Sequence 47, Applicat  | 6.99e-01 |
| 25 | 98  | 43.2 | 20  | 10 | US-08-467- Sequence 230, Applicat | 6.99e-01 |
| 26 | 98  | 43.2 | 21  | 10 | US-08-467- Sequence 231, Applicat | 6.99e-01 |
| 27 | 98  | 43.2 | 23  | 8  | US-08-350- Sequence 223, Applicat | 6.99e-01 |
| 28 | 98  | 43.2 | 23  | 10 | US-08-467- Sequence 225, Applicat | 6.99e-01 |
| 29 | 98  | 43.2 | 26  | 10 | US-08-468- Sequence 84, Applicat  | 6.99e-01 |
| 30 | 98  | 43.2 | 26  | 8  | US-08-350- Sequence 86, Applicat  | 6.99e-01 |
| 31 | 98  | 43.2 | 26  | 7  | US-08-226- Sequence 86, Applicat  | 6.99e-01 |
| 32 | 98  | 43.2 | 26  | 8  | US-08-350- Sequence 83, Applicat  | 6.99e-01 |
| 33 | 98  | 43.2 | 28  | 10 | US-08-467- Sequence 223, Applicat | 6.99e-01 |
| 34 | 98  | 43.2 | 28  | 8  | US-08-350- Sequence 223, Applicat | 6.99e-01 |
| 35 | 98  | 43.2 | 30  | 7  | US-08-226- Sequence 81, Applicat  | 6.99e-01 |
| 36 | 98  | 43.2 | 30  | 10 | 'S-08-467- Sequence 221, Applicat | 6.99e-01 |
| 37 | 98  | 43.2 | 30  | 8  | 'S-08-350- Sequence 79, Applicat  | 6.99e-01 |
| 38 | 98  | 43.2 | 30  | 8  | US-08-350- Sequence 221, Applicat | 6.99e-01 |
| 39 | 98  | 43.2 | 30  | 10 | US-08-467- Sequence 221, Applicat | 6.99e-01 |
| 40 | 98  | 43.2 | 30  | 10 | US-08-467- Sequence 79, Applicat  | 6.99e-01 |
| 41 | 98  | 43.2 | 30  | 10 | US-08-468- Sequence 80, Applicat  | 6.99e-01 |
| 42 | 98  | 43.2 | 90  | 10 | US-08-467- Sequence 64, Applicat  | 6.99e-01 |
| 43 | 98  | 43.2 | 90  | 10 | US-08-467- Sequence 64, Applicat  | 6.99e-01 |
| 44 | 98  | 43.2 | 346 | 3  | 'S-08-350- Sequence 17, Applicat  | 6.99e-01 |
| 45 | 98  | 43.2 | 374 | 10 | US-08-467- Sequence 2, Applicat   | 6.99e-01 |

ALIGNMENTS

RESULT 1  
ID US-08-465-248-36 STANDARD: PRI: 93 AA.  
XX xxxxxx

DE Sequence 36, Application US/08455248  
XX Sequence 36, Application US/08455248  
CC GENERAL INFORMATION:  
CC APPLICANT: Labigne, Agnes  
CC APPLICANT: Sauerbaum, Sebastian  
CC APPLICANT: Ferrero, Richard L.  
CC APPLICANT: Thibierge, Jean-Michel  
CC TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS AGAINST  
CC TITLE OF INVENTION: HELICOBACTER INFECTION, POLYPEPTIDES FOR USE IN THE  
CC TITLE OF INVENTION: COMPOSITIONS, AND NUCLEIC ACID SEQUENCES ENCODING SAID  
CC TITLE OF INVENTION: POLYPEPTIDES  
CC NUMBER OF SEQUENCES: 44  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: FLORENCE BERNARD, Database, Gattolli A  
CC ADDRESSEE: Lander  
CC STREET: 1300 I Street, N.W  
CC CITY: Washington  
CC STATE: D.C.  
CC COUNTRY: USA  
CC ZIP: 20005-3315  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: Patent Release #1.0, Version #1.3C  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/466,248  
CC FILING DATE: 06-JUN-1995  
CC CLASSIFICATION: 435  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 08/447,177  
CC FILING DATE: 19-MAY-1995  
CC CLASSIFICATION: 435  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 08/432,697  
CC FILING DATE: 02-MAY-1995



CC TYPE: amino acid  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: peptide  
CC FRAGMENT TYPE: internal  
SQ SEQUENCE 19 AA: 218 MW: 2058 CN:

Query Match 46.3% Score 105; DB 10; Length 19;  
Best Local Similarity 87.5%; Pred. No. 1.39e-01;  
Matches 14; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 2 KSKMKVTVAFNQFGPN 17  
1: |||||.....  
QY 17 KRRMKVTVAFNQFGPN 32

RESULT 4  
ID US-09-350-225-236 STANDARD: PRT: 19 AA:

XX AC xxxxxx

XX XX

XX DI

XX DE

XX XX

CC Sequence 236, Application US/09350225

CC GENERAL INFORMATION:

CC APPLICANT: Griffith, Irwin J.;

CC APPLICANT: Pollock, Joanne;

CC APPLICANT: Bond, Julian F.;

CC APPLICANT: Garman, Richard D.;

CC APPLICANT: Kuo, Mei-Chang;

CC APPLICANT: Yeung, Siu-mei H.;

CC APPLICANT: Brauer, Andrew;

CC APPLICANT: Exley, Mark A.;

CC APPLICANT: Powers, Steven P.;

CC TITLE OF INVENTION: Allergenic Proteins And Peptides From

CC JAPAN OF INVENTION: Japanese Cedar Pollen

CC NUMBER OF SEQUENCES: 261

CC CORRESPONDENCE ADDRESS:

CC ADDRESSEE: Immunologic Pharmaceutical Corporation, Inc.

CC STREET: 610 Lincoln St

CC CITY: Waltham

CC STATE: MA

CC COUNTRY: USA

CC ZIP: 02154

CC COMPUTER READABLE FORM:

CC MEDIUM TYPE: Floppy disk

CC COMPUTER: IBM PC compatible

CC OPERATING SYSTEM: PC-DOS/MS-DOS

CC SOFTWARE: Patent In Release #1.0, Version #1.25

CC CURRENT APPLICATION DATA:

CC MOLECULE TYPE: peptide  
CC FRAGMENT TYPE: internal  
SQ SEQUENCE 19 AA: 2181 MW: 2058 CN:

Query Match 46.3% Score 105; DB 8; Length 19;  
Best Local Similarity 87.5%; Pred. No. 1.39e-01;  
Matches 14; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 2 KSKMKVTVAFNQFGPN 17  
1: |||||.....  
QY 17 KRRMKVTVAFNQFGPN 32

RESULT 5  
ID US-08-467-006-236 STANDARD: PRT: 19 AA:

XX AC xxxxxx

XX XX

XX DI

XX DE

XX XX

CC Sequence 236, Application US/08467006

CC GENERAL INFORMATION:

CC APPLICANT: Griffith, Irwin J.;

CC APPLICANT: Pollock, Joanne;

CC APPLICANT: Bond, Julian F.;

CC APPLICANT: Garman, Richard D.;

CC APPLICANT: Kuo, Mei-Chang;

CC APPLICANT: Yeung, Siu-mei H.;

CC APPLICANT: Brauer, Andrew;

CC APPLICANT: Exley, Mark A.;

CC APPLICANT: Powers, Steven P.;

CC TITLE OF INVENTION: Allergenic Proteins And Peptides From

CC JAPAN OF INVENTION: Japanese Cedar Pollen

CC NUMBER OF SEQUENCES: 261

CC CORRESPONDENCE ADDRESS:

CC ADDRESSEE: Immunologic Pharmaceutical Corporation, Inc.

CC STREET: 610 Lincoln St

CC CITY: Waltham

CC STATE: MA

CC COUNTRY: USA

CC ZIP: 02154

CC COMPUTER READABLE FORM:

CC MEDIUM TYPE: Floppy disk

CC COMPUTER: IBM PC compatible

CC OPERATING SYSTEM: PC-DOS/MS-DOS

CC SOFTWARE: Patent In Release #1.0, Version #1.25

CC CURRENT APPLICATION DATA:

CC APPLICATION NUMBER: US/08/467,006

CC FILING DATE: June 6, 1995

CC CLASSIFICATION: 424

CC PRIOR APPLICATION DATA:

CC APPLICATION NUMBER: US/350,225

CC FILING DATE: December 6, 1994

CC ATTORNEY/AGENT INFORMATION:

CC NAME: Jane E. Remillard

CC REGISTRATION NUMBER: 38,872

CC REFERENCE/DOCKET NUMBER: 025.6 USD6 (IMI-028CPD6)

CC TELECOMMUNICATION INFORMATION:

CC TELEPHONE: (617) 227-7400

CC TELEFAX: (617) 227-5941

CC INFORMATION FOR SEQ ID NO: 236:

CC SEQUENCE CHARACTERISTICS:

CC LENGTH: 19 amino acids

CC TYPE: amino acid

CC TOPOLOGY: linear

CC MOLECULE TYPE: peptide

CC FRAGMENT TYPE: internal

CC SEQUENCE 19 AA: 2181 MW: 2058 CN:

Query Match 46.3% Score 105; DB 10; Length 19;  
Best Local Similarity 87.5%; Pred. No. 1.39e-01;

Matches 14: Conservative 1: Mismatches 1: Indels 0: Gaps 0:

Db 2 KSKMKVTVAFNQFGPN 17  
1: |||||.....

QY 17 KSKMKVTVAFNQFGPN 32

RESULT 6

ID US-08-467-023-236 STANDARD: PRT: 19 AA.

XX xxxxxx

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

Matches 14: Conservative 1: Mismatches 1: Indels 0: Gaps 0:

Db 2 KSKMKVTVAFNQFGPN 17

1: |||||.....

QY 17 KSKMKVTVAFNQFGPN 32

RESULT 7

ID US-08-467-697-236 STANDARD: PRT: 19 AA.

XX xxxxxx

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

DT XX Sequence 222, Application US/09458940  
 DE XX Sequence 222, Application US/09458940  
 CC XX GENERAL INFORMATION:  
 CC APPLICANT: Griffith, Irwin J.;  
 CC APPLICANT: Pollock, Joanne;  
 CC APPLICANT: Bond, Julian F.;  
 CC APPLICANT: Garman, Richard D;  
 CC APPLICANT: Kuo, Mei-Chang;  
 CC APPLICANT: Yeung, Siu-wei H.;  
 CC APPLICANT: Brauer, Andrew;  
 CC APPLICANT: Exley, Mark A.;  
 CC APPLICANT: Powers, Steven P.  
 CC TITLE OF INVENTION: Allergenic Proteins And Peptides From  
 CC TITLE OF INVENTION: Japanese Cedar Pollen  
 CC NUMBER OF SEQUENCES: 26;  
 CC CORRESPONDENCE ADDRESS:  
 CC ADDRESSEE: Immunologic Pharmaceutical Corporation, Inc.  
 CC STREET: 610 Lincoln St  
 CC CITY: Waltham  
 CC STATE: MA  
 CC COUNTRY: USA  
 CC ZIP: 02154  
 CC COMPUTER READABLE FORM:  
 CC MEDIUM TYPE: Floppy disk  
 CC COMPUTER: IBM PC compatible  
 CC OPERATING SYSTEM: PC-DOS/MS-DOS  
 CC SOFTWARE: Patent In Release #1.0, Version #1.25  
 CC CURRENT APPLICATION DATA:  
 CC APPLICATION NUMBER: US/08/458,940  
 CC FILING DATE:  
 CC CLASSIFICATION: 424  
 CC PRIOR APPLICATION DATA:  
 CC APPLICATION NUMBER: 08/350,225  
 CC FILING DATE: December 6, 1994  
 CC APPLICATION NUMBER: 08/226,248  
 CC FILING DATE: April 8, 1994  
 CC APPLICATION NUMBER: 07/938,990  
 CC FILING DATE: September 1, 1992  
 CC APPLICATION NUMBER: PCT/US93/00139  
 CC FILING DATE: January 15, 1993  
 CC ATTORNEY/AGENT INFORMATION:  
 CC NAME: Darlene A. Vanstone  
 CC REGISTRATION NUMBER: 35,729  
 CC REFERENCE/DOCKET NUMBER: 025.6 US (IM1-028CP2)  
 CC TELECOMMUNICATION INFORMATION:  
 CC TELEPHONE: (617) 466-6000  
 CC TELEFAX: (617) 466-6040  
 CC INFORMATION FOR SEQ ID NO: 222:  
 CC SEQUENCE CHARACTERISTICS:  
 CC LENGTH: 29 amino acids  
 CC TYPE: amino acid  
 CC TOPOLOGY: linear  
 CC MOLECULE TYPE: peptide  
 CC FRAGMENT TYPE: internal  
 CC SEQUENCE 29 AA: 3313 MW: 4103 CN:  
 SQ  
 Query Match 46.3%; Score 105; DB 10; Length 29;  
 Best Local Similarity 87.5%; Pred. No. 1,39e-01;  
 Matches 14; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 Db 1 KSKMKVTVAFNQFGPN 16  
 QY 17 KRMKVVTVAFNQFGPN 32  
 RESULT 9  
 ID US-08-350-225-222 STANDARD: PRT: 29 AA.  
 XX  
 AC xxxxxx  
 XX  
 DT XX

DT XX Sequence 222, Application US/08350225  
 DE XX Sequence 222, Application US/08350225  
 CC XX GENERAL INFORMATION:  
 CC APPLICANT: Griffith, Irwin J.;  
 CC APPLICANT: Pollock, Joanne;  
 CC APPLICANT: Bond, Julian F.;  
 CC APPLICANT: Garman, Richard D;  
 CC APPLICANT: Kuo, Mei-Chang;  
 CC APPLICANT: Yeung, Siu-wei H.;  
 CC APPLICANT: Brauer, Andrew;  
 CC APPLICANT: Exley, Mark A.;  
 CC APPLICANT: Powers, Steven P.  
 CC TITLE OF INVENTION: Allergenic Proteins And Peptides From  
 CC TITLE OF INVENTION: Japanese Cedar Pollen  
 CC NUMBER OF SEQUENCES: 26;  
 CC CORRESPONDENCE ADDRESS:  
 CC ADDRESSEE: Immunologic Pharmaceutical Corporation, Inc.  
 CC STREET: 610 Lincoln St  
 CC CITY: Waltham  
 CC STATE: MA  
 CC COUNTRY: USA  
 CC ZIP: 02154  
 CC COMPUTER READABLE FORM:  
 CC MEDIUM TYPE: Floppy disk  
 CC COMPUTER: IBM PC compatible  
 CC OPERATING SYSTEM: PC-DOS/MS-DOS  
 CC SOFTWARE: Patent In Release #1.0, Version #1.25  
 CC CURRENT APPLICATION DATA:  
 CC APPLICATION NUMBER: US/08/350,225  
 CC FILING DATE: December 6, 1994  
 CC CLASSIFICATION: 424  
 CC PRIOR APPLICATION DATA:  
 CC APPLICATION NUMBER: 08/226,248  
 CC FILING DATE: April 8, 1994  
 CC APPLICATION NUMBER: 07/938,990  
 CC FILING DATE: September 1, 1992  
 CC APPLICATION NUMBER: PCT/US93/00139  
 CC FILING DATE: January 15, 1993  
 CC ATTORNEY/AGENT INFORMATION:  
 CC NAME: Darlene A. Vanstone  
 CC REGISTRATION NUMBER: 35,729  
 CC REFERENCE/DOCKET NUMBER: 025.6 US (IM1-028CP2)  
 CC TELECOMMUNICATION INFORMATION:  
 CC TELEPHONE: (617) 466-6000  
 CC TELEFAX: (617) 466-6040  
 CC INFORMATION FOR SEQ ID NO: 222:  
 CC SEQUENCE CHARACTERISTICS:  
 CC LENGTH: 29 amino acids  
 CC TYPE: amino acid  
 CC TOPOLOGY: linear  
 CC MOLECULE TYPE: peptide  
 CC FRAGMENT TYPE: internal  
 CC SEQUENCE 29 AA: 3313 MW: 4103 CN:  
 SQ  
 Query Match 46.3%; Score 105; DB 8; Length 29;  
 Best Local Similarity 87.5%; Pred. No. 1,39e-01;  
 Matches 14; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 Db 1 KSKMKVTVAFNQFGPN 16  
 QY 17 KRMKVVTVAFNQFGPN 32  
 RESULT 10  
 ID US-08-467-023-222 STANDARD: PRT: 29 AA.  
 XX  
 AC xxxxxx  
 XX  
 DT XX



```
CC STREET: 610 Lincoln St  
CC CITY: Waltham  
CC STATE: MA  
CC COUNTRY: USA  
CC ZIP: 02154  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: Patent Release 2.0, Version 2.25  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/350,225  
CC FILING DATE: December 6, 1994  
CC CLASSIFICATION: 424  
CC PRIORITY APPLICATION DATA:  
CC APPLICATION NUMBER: 08/225,248  
CC FILING DATE: April 8, 1994  
CC APPLICATION NUMBER: 07/938,990  
CC FILING DATE: September 1, 1992  
CC APPLICATION NUMBER: PCI/US93/00139  
CC FILING DATE: January 15, 1993  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Darlene A. Vanstone  
CC REGISTRATION NUMBER: 95,729  
CC REFERENCE/DOCKET NUMBER: 025.6 US (CI-028CP2)  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (617) 466-8000  
CC TELEFAX: (617) 466-6040  
CC INFORMATION FOR SEQ ID NO: 228:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 26 amino acids  
CC TYPE: amino acid  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: peptide  
CC FRAGMENT TYPE: internal  
CC  
SQ SEQUENCE 26 AA; 2956 MW; 3132 CN;  
  
Query Match 45.4%; Score 193; DB 8; Length 26;  
Best Local Similarity 93.3%; Prd No. 2,21e-03;  
Matches 14; Conservative 0; Mismatches 1; Indels 0;  
  
Db 1 RSMKVTVAFNQFGPN 15  
I | | | | | | | |  
QY 18 PRMKVTVAFNQFGPN 32  
  
RESULT 14  
ID US-08-350-225-229 STANDARD: PRT: 26 AA.  
AC xxxxxx  
XX  
XX  
XX  
XX  
XX DE  
XX  
XX  
Sequence 225, Application: US/08350225  
Sequence 229, Application: US/08350225  
GENERAL INFORMATION:  
APPLICANT: Griffith, Irwin S.;  
APPLICANT: Pollock, Joanne;  
APPLICANT: Bond, Julian F.;  
APPLICANT: Garman, Richard D.;  
APPLICANT: Kuo, Mei-Chang;  
APPLICANT: Yeung, Siu-mei H.;  
APPLICANT: Brauer, Andrew;  
APPLICANT: Exley, Mark A.;  
APPLICANT: Powers, Steven P.  
TITLE OF INVENTION: Allergenic Proteins And Peptides From Japanese Cedar Pollen  
NUMBER OF SEQUENCES: 261  
CORRESPONDENCE ADDRESS:  
ADDRESS: Immunologic Pharmaceutical Corporation, Inc.  
STREET: 610 Lincoln St  
CITY: Waltham
```

```

CC ZIP: 02154
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS
CC SOFTWARE: Patentin Release # 10, Version # 1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/467,023
CC FILING DATE: June 6, 1996
CC CLASSIFICATION: 424
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/150,229
CC FILING DATE: December 6, 1994
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Jane E. Remillard
CC REGISTRATION NUMBER: 56,72
CC REFERENCE/DOCKET NUMBER: 25,152 US/4 (IM-0280P12)
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (617) 227-7
CC TELEFAX: (617) 227-594
CC INFORMATION FOR SEQ ID NO: 1:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 26 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC FRAGMENT TYPE: Internal
CC SEQUENCE: 26 AA: 2964 MW: 1481 CN:

Query Match: 45.4%; Score 163; DB 10; Length 250;
Best Local Similarity 93.3%; Pred.No. 2 21e-01;
Matches 14; Conservative 0; Mismatches 1; Indels

Db 1 RSMKVTVAVFNGFGPN 15
| | | | | | | | |
Oy 18 RSMKVTVAVFNGFGPN 32

Search completed: Mon Jun 19 16:26:53 2000
Job time : 19 secs.

```



\*\*\*\*\*

[illegible]

Release 3.1A John F. Collins, BioComputing Research Unit  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

```

mpsrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on:      Mon Jun 19 16:12:03 2000:      Wapser time 6.47 Seconds
                                         233.465 Million cell updates/sec
Tabular output not generated.

```

```

Title:
Description:
Perfect Score:
Sequence:

```

Scoring table: PAM 150  
Gap 11

Searched: 142080 seqs, 47172406 residues

Post-processing: Minimum Match 08  
Listing first 45 summaries

```
Database:
pir52
1:pir1 2:pir2 3:pir3 4:pir4
```

Statistics: Mean 34.452; Variance 60.396; scale 0.570

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

| Result<br>No. | Score | Query |     | Length | DB     | ID | Description                        | Prod     | No. |
|---------------|-------|-------|-----|--------|--------|----|------------------------------------|----------|-----|
|               |       | Match | %   |        |        |    |                                    |          |     |
| 1             | 114   | 50.2  | 100 | 2      | JF5180 |    | heat shock protein gp              | 1.72e-06 |     |
| 2             | 114   | 50.2  | 100 | 2      | SH5556 |    | 10k - cel. antigen -               | 1.72e-06 |     |
| 3             | 114   | 50.2  | 100 | 1      | BUYVBA |    | chaperonin groES - My              | 1.72e-06 |     |
| 4             | 114   | 50.2  | 169 | 2      | S72818 |    | heat shock protein ch              | 1.72e-06 |     |
| 5             | 105   | 46.3  | 102 | 2      | S75685 |    | chaperonin groES - St              | 5.50e-05 |     |
| 6             | 103   | 45.4  | 102 | 2      | A41325 |    | heat shock protein 18              | 1.43e-04 |     |
| 7             | 98    | 43.2  | 374 | 2      | JC2124 |    | major allergen Cyt 1               | 1.00e-03 |     |
| 8             | 98    | 43.2  | 374 | 2      | JC2123 |    | major allergen Cyt 1               | 1.00e-03 |     |
| 9             | 88    | 38.8  | 242 | 2      | EY1621 |    | ERCC-like excision 1               | 4.27e-02 |     |
| 10            | 84    | 37.0  | 455 | 2      | TC856  |    | pectate lyase (EC 4.2              | 1.80e-01 |     |
| 11            | 80    | 35.2  | 103 | 1      | BYVQCS |    | chaperonin groES - Sy              | 7.31e-01 |     |
| 12            | 80    | 35.2  | 103 | 2      | A36711 |    | groES protein - Sync               | 7.31e-01 |     |
| 13            | 80    | 35.2  | 310 | 2      | TI1320 |    | NADH dehydrogenase (u              | 7.31e-01 |     |
| 14            | 80    | 35.2  | 328 | 2      | TI1938 |    | NADH dehydrogenase (u              | 7.31e-01 |     |
| 15            | 79    | 34.8  | 100 | 1      | BYMY7B |    | chaperonin groES - My              | 1.03e-00 |     |
| 16            | 79    | 34.8  | 869 | 1      | A47257 |    | 1-phosphatidylinositol             | 1.03e-00 |     |
| 17            | 78    | 34.4  | 103 | 2      | T06830 |    | chaperonin groES - Cy              | 1.45e+00 |     |
| 18            | 78    | 34.4  | 230 | 1      | F53610 |    | H <sup>+</sup> -transporting ATP s | 1.45e+00 |     |
| 19            | 78    | 34.4  | 330 | 2      | C58931 |    | NADH dehydrogenase (u              | 1.45e+00 |     |
| 20            | 76    | 33.5  | 96  | 2      | C49377 |    | heat shock protein hs              | 2.84e+00 |     |
| 21            | 76    | 33.5  | 325 | 1      | B40358 |    | NADH dehydrogenase (u              | 2.84e+00 |     |
| 22            | 76    | 33.5  | 325 | 1      | DNMT11 |    | NADH dehydrogenase (u              | 2.84e+00 |     |
| 23            | 76    | 33.5  | 325 | 1      | S49576 |    | NADH dehydrogenase (u              | 2.84e+00 |     |

|    |    |      |      |   |        |                        |          |
|----|----|------|------|---|--------|------------------------|----------|
| 24 | 76 | 33.5 | 331  | 1 | DNOB1  | NADH dehydrogenase (u  | 1.94E+00 |
| 25 | 78 | 33.0 | 94   | 2 | A9855  | heat shock protein (u  | 3.3E+00  |
| 26 | 78 | 33.0 | 94   | 2 | A1884  | heat shock protein (c  | 3.3E+00  |
| 27 | 75 | 33.0 | 96   | 2 | F152   | hypothetical protein   | 3.9E+00  |
| 28 | 75 | 33.0 | 225  | 2 | F1090  | hypothetical protein   | 3.9E+00  |
| 29 | 74 | 32.6 | 94   | 2 | C2479  | heat shock protein (G  | 5.51E+00 |
| 30 | 74 | 32.6 | 894  | 2 | C2479  | cell wall-associated   | 5.51E+00 |
| 31 | 74 | 32.6 | 1006 | 2 | F59730 | kinase-defective Eph-  | 5.51E+00 |
| 32 | 73 | 32.2 | 88   | 2 | PC1238 | heat shock protein (B- | 7.43E+00 |
| 33 | 73 | 32.2 | 94   | 2 | S2613  | chaperonin, glucose (H | 7.43E+00 |
| 34 | 73 | 32.2 | 129  | 2 | T15050 | NADH dehydrogenase (u  | 7.63E+00 |
| 35 | 73 | 32.2 | 404  | 2 | T25556 | pectate lyase (EC 4.2  | 7.63E+00 |
| 36 | 73 | 32.2 | 517  | 2 | TG299  | cytochrome P450 (anol  | 7.63E+00 |
| 37 | 72 | 31.7 | 204  | 2 | F5853  | deoxycyclidine triphos | 1.05E+01 |
| 38 | 72 | 31.7 | 501  | 2 | H5890  | hypothetical protein   | 1.05E+01 |
| 39 | 72 | 31.7 | 404  | 2 | S12209 | pectate lyase (EC 4.2  | 1.05E+01 |
| 40 | 72 | 31.7 | 434  | 2 | S29612 | pectate lyase (EC 4.2  | 1.05E+01 |
| 41 | 72 | 31.7 | 438  | 2 | S43335 | pectate lyase (EC 4.2  | 1.05E+01 |
| 42 | 72 | 31.7 | 1213 | 2 | VWU    | von Willebrand factor  | 1.05E+01 |
| 43 | 71 | 31.3 | 121  | 1 | 158091 | NF-kappa B complex su  | 1.45E+01 |
| 44 | 71 | 31.3 | 326  | 2 | S5931  | NADH dehydrogenase (u  | 1.45E+01 |
| 45 | 71 | 31.3 | 558  | 2 | A42023 | transcription activat  | 1.45E+01 |

## ALIGNMENTS

```

RESULT      1
ENTRY
  TITLE      S25180      #type complete
  ALTERNATE_NAMES      heat shock protein groES - Mycobacterium leprae
  ORGANISM      10K chaperonin
  DATE      #normal_name Mycobacterium leprae
  #Feb-1994 #sequence_revision 03-Feb-1994 #text_change
  26-Aug-1999
ACCESSIONS      S25180: S72997
REFERENCE
  #authors      de Wit, T.F.R.; Bekkerie, S.; Osland, A.; Mike, J.L.L.; Hermans,
  P.M.; van Soestijnen, D.; Drijfhout, J.; Schepelijn, H.;
  Janson, A.A.M.; Thore, J.E.R.
  #journal      Mol. Microbiol. (1992) 6:1955-2007
  #title      Mycobacteria contain two groEL genes: the second
  Mycobacterium leprae groEL gene is arranged in an operon
  with groES.
  #cross-references      MUID:92374850
  #accession      S25180
  #molecule_type      DNA
  #residues      1-100 #label SM1
  #cross-references      EMBL:Z11665
  #note      the authors found the following single amino
  residue as the
REFERENCE
  S72580
  #authors      Smith, D.R.; Robinson, K.
  #submission      Submitted to the EMBL Data Library, November: 1993
  #description      Mycobacterium leprae cosmid B239.
  #accession      S72997
  #status      preliminary
  #molecule_type      DNA
  #residues      1-100 #label SM1
  #cross-references      EMBL:U00026; NID:q467102; PID:AAA17311.1;
  PID:q467139

```

94 EEYLILSARDVLAVSK 100  
17

QY 1 EEYLILSARDVLAVSK 17

```

RESULTS
ENTRY 4
TITLE 10K T-cell antigen - Mycobacterium leprae
ORGANISM Mycobacterium leprae
DATE 30-Jan-1992 #sequence_revision 30-Jun-1992 #text_change 12-Sep-1997
ACCESSIONS J03556
REFERENCE J03556
AUTHORS Mehra, V.; Bloom, B.R.; Barardi, A.C.; Grisso, G.L.; Sieling, P.A.; Alland, D.; Convit, C.; Fan, X.; Hunter, S.W.; Brennan, P.J.; Roa, T.H.; Modlin, R.L.
J. Exp. Med. (1992) 175:275-284
A major T cell antigen of Mycobacterium leprae is a 10-KD heat-shock cognate protein.
CROSS-REFERENCES NCID:92113469
ACCESSION J03556
STATUS nucleic acid sequence not shown
MOLECULE_TYPE DNA
RESIDUES 1-100 #label MEH
CROSS-REFERENCES EMBL:X59413
COMMENT This protein is a heat-shock protein.
CLASSIFICATION superfamily chaperonin groES
SUMMARY length 100 molecular-weight 10768 #checksum 7371
Query Match 50.2% Score 114 DB 2 Length 100
Best Local Similarity 100.0% Pred. No. 1,720-06;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
DE 84 EBYLLSARDVLAVVSK 100
QY 1 EBYLLSARDVLAVVSK 17
RESULTS 3
ENTRY 3
TITLE Chaperonin groES - Mycobacterium tuberculosis
ALTERNATE_NAMES 10K antigen; BCG-a homolog; Cpn10; heat shock protein 10K (hsp10); immunogenic protein BCG-a
ORGANISM Mycobacterium tuberculosis
DATE 31-Dec-1990 #sequence_revision 31-Dec-1990 #text_change 16-Jul-1995
ACCESSIONS S01381; A37166; S02727; A47292; G70747; A46481; S18040
REFERENCE S01381
AUTHORS Baird, P.N.; Hall, L.M.C.; Coates, A.R.M.
Nucleic Acids Res. (1986) 14:9347
A major antigen from Mycobacterium tuberculosis which is homologous to the heat shock proteins groES from E. coli and the hspA gene product of Escherichia coli.
CROSS-REFERENCES NCID:9016484
ACCESSION S01381
MOLECULE_TYPE DNA
RESIDUES 1-100 #label BAI
CROSS-REFERENCES EMBL:X12598; NID:q44551; PID:q581352
REFERENCE A37166
AUTHORS Baird, P.N.; Hall, L.M.C.; Coates, A.R.M.
J. Gen. Microbiol. (1989) 135:931-939
Cloning and sequence analysis of the 10 kDa antigen gene of Mycobacterium tuberculosis.
CROSS-REFERENCES NCID:90095443
ACCESSION A37166
STATUS preliminary
MOLECULE_TYPE DNA
RESIDUES 1-2-100 #label BA2
CROSS-REFERENCES G5:M25258; G3:X12598
REFERENCE S02727
AUTHORS Shinnick, T.M.; Plikaytis, B.B.; Hyshe, A.D.; van Landingham, R.M.; Walker, L.L.
Nucleic Acids Res. (1989) 17:1254
The Mycobacterium tuberculosis BCG-a protein has homology with the Escherichia coli GroES protein.
CROSS-REFERENCES NCID:89150258

```

```

ACCESSION S02727
MOLECULE_TYPE DNA
RESIDUES 1-100 #label SHI
CROSS-REFERENCES EMBL:X13739; NID:q44571; PID:CAA32303; PID:q581460
REFERENCE A47292
AUTHORS Kong, T.H.; Coates, A.R.; Butcher, P.D.; Hickman, C.; Shinnick, T.M.
Proc. Natl. Acad. Sci. U.S.A. (1994) 90:2639-2712
Mycobacterium tuberculosis expresses two chaperonin homologs.
CROSS-REFERENCES NCID:93219332
ACCESSION A47292
STATUS preliminary
MOLECULE_TYPE DNA
RESIDUES 1-100 #label KON
CROSS-REFERENCES EMBL:X60350; NID:q44599; PID:CAA42908; PID:q581363
SEQUENCE extracted from NCBI backbone (NCBI:128605; NCBI:P:128607)
REFERENCE A70500
AUTHORS Cole, S.J.; Brosch, R.; Parkhill, J.; Garnier, J.; Churcher, C.; Harris, D.; Gordon, S.V.; Eigmeier, K.; Gas, S.; Barry, III, C.E.; Tekala, F.; Badcock, K.; Basham, D.; Brown, D.; Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Harlin, N.; Holroyd, S.; Hornsby, T.; Jagels, K.; Krogh, A.; McLean, J.; Miller, S.; Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Squires, R.; Sulston, J.; Taylor, K.; Whitehead, S.; Barrett, B.G.
Nature (1998) 393:537-544
Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence.
CROSS-REFERENCES NCID:98295987
ACCESSION G70737
STATUS nucleic acid sequence not shown; translation not shown
MOLECULE_TYPE DNA
RESIDUES 1-100 #label GCL
CROSS-REFERENCES GB:Z77165; DB:A123456; NID:q1261609; PID:CAAC01051; PID:q55190; PID:q1449369
EXPERIMENTAL_SOURCE strain 1499
REFERENCE A4648
AUTHORS Barnes, P.F.; Mehra, V.; Kivett, P.; Fong, S.J.; Brennan, P.J.; Voegtline, M.S.; Minden, P.; Hopfinger, R.A.; Bloom, B.R.; Modlin, R.L.
J. Immunol. (1992) 148:1845-1849
Immunoreactivity of a 10-kDa antigen of Mycobacterium tuberculosis.
CROSS-REFERENCES NCID:92171547
ACCESSION A4648
STATUS preliminary
MOLECULE_TYPE DNA
RESIDUES 1-100 #label HAK
SEQUENCE extracted from N261 backbone (NCBI:P:87128)
GENETICS
GENE groES; cpn10
START_CODON GTG
COMPLEX functional chaperonin includes 14 chains of groEL and 7 of groES
FUNCTION mediates protein folding and renaturation
CLASSIFICATION superfamily chaperonin groES
KEYWORDS ATP; heat shock; molecular chaperone; stress-induced protein
FEATURE
2-100 #product chp; renin groES #status experimental #label MAT
SUMMARY #length 100 #molecular-weight 10804 #checksum 7634
Query Match 50.2% Score 114 DB 1 Length 100
Best Local Similarity 100.0% Pred. No. 1,720-06;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
DB 84 EBYLLSARDVLAVVSK 100
1 EBYLLSARDVLAVVSK 17

```

QY 1 BEYLILSARDVLAVVSK 17

RESULT 4

ENTRY  
TITLE heat shock protein chpA - Mycobacterium leprae  
ALTERNATE\_NAMES chaperonin, 10k; protein B1620.C3\_227  
ORGANISM Mycobacterium leprae  
DATE 19-Mar-1997 #sequence\_revision 25-Apr-1997 #text\_change 26-Aug-1999

ACCESSIONS S72818  
REFERENCE S72550  
#authors Smith, D.P.; Robinson, K  
#submissions submitted to the FMBL Data Library, November 1993  
#description Mycobacterium leprae cosmid B1620.  
#accession S72818  
#status preliminary  
#molecule\_type DNA  
#residues 1-169 #label SW2  
#cross-references EMBL:000015; NID:g466931; PIDN:AA03227.1; PID:g466935

GENETICS

#gene chpA  
#superfamily chaperonin groES  
#keywords heat shock; molecular chaperone; stress-induced protein  
#summary length 169 #molecular\_weight 18097 #checksum 6011

Query Match 50.2%; Score 114; DB 2; Length 169;  
Best Local Similarity 100.0%; Pred. No. 1.72e-06;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 153 BEYLILSARDVLAVVSK 169

QY 1 BEYLILSARDVLAVVSK 17

RESULT 5

ENTRY  
TITLE Chaperonin groES - Streptomyces coelicolor  
ALTERNATE\_NAMES chaperonin, 10k; protein B1620.C3\_227  
ORGANISM Streptomyces coelicolor  
DATE 06-Jan-1995 #sequence\_revision 06-Jan-1995 #text\_change 26-Aug-1999

ACCESSIONS S37565  
REFERENCE S37564  
#authors Ducheene, A.M.; Kieser, K.; Hopwood, D.; Thompson, C.; Mazodier, P.  
#submissions submitted to the FMBL Data Library, September 1993  
#description Molecular characterization of two groEL genes in Streptomyces coelicolor A1(2).  
#accession S37565  
#status preliminary  
#molecule\_type DNA  
#residues 1-132 #label BJC  
#cross-references EMBL:X75206; NID:g406595; PIDN:CAA51018.1; PID:g809756

GENETICS

#gene groES  
#start\_codon 31G  
#superfamily chaperonin groES  
#keywords heat shock; molecular chaperone; stress-induced protein  
#summary length 102 #molecular\_weight 10946 #checksum 4809

Query Match 46.3%; Score 105; DB 2; Length 102;  
Best Local Similarity 82.4%; Pred. No. 5.50e-05;  
Matches 14; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

DB 86 BEYLILSARDVLAVVSK 102

QY 1 BEYLILSARDVLAVVSK 17

RESULT 6

ENTRY A41325 #type complete

TITLE heat shock protein 18 - Streptomyces albus  
ALTERNATE\_NAMES heat shock protein; groES homolog  
ORGANISM Streptomyces albus  
DATE 17-Jul-1992 #sequence\_revision 17-Jul-1992 #text\_change 12-Sep-1997

ACCESSIONS A41325  
REFERENCE A41325  
#authors Mazodier, P.; Sugliemi, G.; Davies, J.; Thompson, C.  
#journal J. Bacteriol. (1992) 173:7392-7386  
#title Characterization of the groEL-like genes in Streptomyces albus.  
#cross-references MUID:92041639

#accession A41325  
#molecule\_type DNA  
#residues 1-102 #label MAZ  
#cross-references GB:M76657

GENETICS

#start\_codon GTG  
#superfamily chaperonin groES  
#classification length 102 #molecular\_weight 10955 #checksum 4643  
#summary

Query Match 45.4%; Score 103; DB 2; Length 102;  
Best Local Similarity 76.5%; Pred. No. 2.43e-04;  
Matches 13; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

DB 86 BEYLILSARDVLAVVSK 102

QY 1 BEYLILSARDVLAVVSK 17

RESULT 7

ENTRY JC2124 #type complete  
TITLE major allergen Cry j I precursor (clone pCC1-15) - Japanese cedar  
ORGANISM Cryptomeria japonica  
DATE 14-Jul-1994 #sequence\_revision 14-Jul-1994 #text\_change 26-Aug-1999

ACCESSIONS JC2124  
REFERENCE JC2123  
#authors Sone, T.; Komatsu, N.; Shimizu, K.; Kusakabe, T.; Marikubo, K.; Kino, K.  
#journal Biochem. Biophys. Res. Commun. (1994) 199:619-625  
#title Cloning and sequencing of cDNA coding for Cry j I, a major allergen of Japanese cedar pollen.  
#cross-references MUID:94183234

#accession JC2124  
#molecule\_type mRNA  
#residues 1-374 #label SON  
#cross-references GB:025545; NID:g49443; PIDN:049443; PIDN:049443  
#experimental\_source pollen  
#note the authors described carbohydrate binding site for residue 279

CLASSIFICATION #superfamily peptidyl lyase L4T5  
KEYWORDS glycoprotein; pollen

FEATURE  
1-21  
22-374  
158-159,293,354  
#domain signal sequence #status predicted #label SIG  
#product major allergen Cry j I (clone pCC1-15) #status predicted #label MAZ  
#binding site carbohydrate (Asn) (covalent) #status predicted

SUMMARY length 374 #molecular\_weight 40702 #checksum 3692

Query Match 43.2%; Score 98; DB 2; Length 374;  
Best Local Similarity 86.7%; Pred. No. 1.00e-03;  
Matches 13; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

DB 232 KSMKVTVAFOGPN 246

QY 18 RMKVTVAFOGPN 32

RESULT 8

ENTRY JC2123 #type complete

```

TITLE      major allergen Cry 1 precursor (clone pCCT-2-2) - Japanese cedar
ORGANISM   #formal_name Cryptomeria japonica #common_name Japanese cedar
DATE       14-Jul-1994 #sequence_revision 14-Jul-1994 #text_change
ACCESSIONS J02123: PC2065
REFERENCE   J02123: PC2065
AUTHORS    Sone, T.; Komiya, N.; Shimizu, K.; Kusakabe, T.; Morikubo, K.; Kino, K.
JOURNAL    Biochem. Biophys. Res. Commun. (1994) 199:619-625
TITLE      Cloning and sequencing of cDNA coding for Cry 1, a major allergen of Japanese cedar pollen.
CROSS-REFERENCES MIM:9418324
ACCESSION  J02123
MOLECULE_TYPE mRNA
RESIDUES   1-374 #label SN
CROSS-REFERENCES GB:026544; NID:q493631; PID:d1006066; PID:q493632
EXPERIMENTAL_SOURCE pollen
ACCESSION  PC2065
MOLECULE_TYPE protein
RESIDUES   22-53:58-81;219-232;236-258;299-307;346-372 #label S02
NOTE       the authors described carbohydrate binding site for residue 279
CLASSIFICATION #superfamily pectate lyase LA759
KEYWORDS    glycoprotein; pollen
FEATURE     1-2:
22-374      #domain signal sequence #status predicted #label SIG\
            #product major allergen Cry 1 (clone pCCT-2-2) #status
            predicted #label WAT\
158-191,293,354 #binding_site carbohydrate (Asn) (covalent) #status
            Predicted
SUMMARY     #length 374 #molecular_weight 40545 #checksum 2920
Query Match 43.28; Score 98; DB 2; Length 374;
Best Local Similarity 86.78; Pred. No. 1,00e-03;
Matches 13; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
DB 232 KSKVTVAFTVAFNPGN 246
QY 18 PRMKTVAFTVAFNPGN 12
RESULT 9
ENTRY   E71621 #type complete
TITLE   ERCC-like excision repair protein PFB01604 - malaria parasite (Plasmodium falciparum)
ORGANISM #formal_name Plasmodium falciparum
DATE     13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change
ACCESSIONS E71621
REFERENCE   A71621
AUTHORS    Gardner, M.J.; Tetlow, B.; Garavito, R.M.; Cummings, L.M.; Aravind, L.; Koolin, E.V.; Shalimov, S.; Mason, J.; Yu, K.; Fujii, C.; Pederson, J.; Shen, K.; Jing, J.; Aston, C.; Lal, Z.; Schwartz, G.C.; Pertea, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, R.O.; Fraser, C.M.; Adams, M.D.; Venter, J.C.; Hoffman, S.D. Science (1998) 282:1126-1132
TITLE     Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
CROSS-REFERENCES MIM:99021743
ACCESSION  E71621
STATUS     preliminary; nucleic acid sequence not shown; translation not shown
MOLECULE_TYPE DNA
RESIDUES   1-242 #label GAR
CROSS-REFERENCES GB:AS001377; GB:AS001362; NID:q384511; PID:q384512; #GR:PFB01604
EXPERIMENTAL_SOURCE clone 3D7
GENETICS   #molecular_weight 28257 #checksum 9994
SUMMARY     #length 242 #molecular_weight 28257 #checksum 9994

```

```

Query Match 38.88; Score 88; DB 2; Length 242;
Best Local Similarity 40.88; Pred. No. 4,27e-02;
Matches 13; Conservative 8; Mismatches 11; Indels 0; Gaps 0;
DB 45 EOYLITSLROKLPVKKRKRVYKFN:IPD 76
QY 1 EYLILSARDVLAVSKRRMKVTVAFNCGPN 32
RESULT 10
ENTRY   T00856 #type complete
TITLE   pectate lyase (EC 4.2.2.2) T20F6.14 - Arabidopsis thaliana
ORGANISM #formal_name Arabidopsis thaliana #common_name Mustard; cross
DATE     12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change
ACCESSIONS T00856
REFERENCE   Z14206
AUTHORS    Rounsley, S.D.; Lin, X.; Keichum, K.A.; Crosby, M.; Brandon, R.C.; Sykes, S.M.; Kaul, S.; Mason, J.M.; Kerlavage, A.R.; Adams, M.D.; Somerville, C.R.; Venter, J.C.
SUBMISSION submitted to the C330 Data Library, March 1998
DESCRIPTION Arabidopsis thaliana chromosome 11 BAC T20F6 genomic sequence.
ACCESSION  T00856
STATUS     translated from GB/EMBL/DBE;
MOLECULE_TYPE DNA
RESIDUES   1-455 #label RCU
CROSS-REFERENCES EMBL:AC021111; NID:q4947056; PID:AA0054511; PID:q2944
EXPERIMENTAL_SOURCE cultivar Columbia
GENETICS   #map_position 2
            #introns 66/2; 295/3; 376/3
            #cote T20F6.14
CLASSIFICATION #superfamily pectate lyase LA159
KEYWORDS     carbon-oxygen lyase
SUMMARY      #length 455 #molecular_weight 51257 #checksum 411
Query Match 37.08; Score 84; DB 2; Length 455;
Best Local Similarity 36.78; Pred. No. 1,8e-01;
Matches 11; Conservative 11; Mismatches 6; Indels 2; Gaps 2;
DB 294 DEVLPGGTNK-DVIDKK-MQITVAFNHFQ 321
QY 1 EYLILSARDVLAVSKRRMKVTVAFNCGP 30
RESULT 11
ENTRY   E09608 #type complete
TITLE   Chaperonin (Cpn) - Synecococcus sp. (Strain 6805-61)
ALTERNATE_NAMES Heat shock protein, hsp (hsp8)
ORGANISM #formal_name Synecococcus sp.
DATE     30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change
ACCESSIONS S02836; S09619
REFERENCE   S0286
AUTHORS    Cozens, A.L.; Walker, J.E.
JOURNAL    J. Mol. Biol. (1987) 194:359-383
TITLE     The organization and sequence of the genes for ATP synthase subunits in the cyanobacterium Synecococcus 6805. Support for an endosymbiotic origin of chloroplasts.
CROSS-REFERENCES MIM:87311713
ACCESSION  S10636
MOLECULE_TYPE DNA
RESIDUES   1-103 #label CO2
CROSS-REFERENCES EMBL:X05925; NID:q48021; PID:CAA293611; PID:q48023
COMPLEX    functional chaperonin includes 14 chains of groEL and 7 of groES
FUNCTION    mediates protein folding and renaturation.
CLASSIFICATION #superfamily chaperonin groES

```

KEYWORDS  
 FEATURE  
 2-97  
 SUMMARY

ATP: heat shock: molecular chaperone; stress-induced protein  
 \*product chaperonin groES \*status predicted \*label MAT  
 \*length: 103 \*molecular-weight 10811 \*checksum 6496

Query Match 35.2% Score 80; DB 1; Length 103;  
 Best Local Similarity 50.0%; Pred. No. 7.31e-01;  
 Matches 8; Conservative 7; Mismatches 1; Indels 0; Gaps 0;

DB 88 DQVLLSEKDLAVVA 103  
 QY 1 EYLLSARDVLAVVS 16

RESULT 12  
 ENTRY  
 TITLE  
 ORGANISM  
 DATE

A36721 \*type complete  
 groES protein - *Synechococcus* sp. (strain PCC 7942)  
 \*formal\_name *Synechococcus* sp.  
 12-Apr-1991 \*sequence\_revision 12-Apr-1991 \*text\_change  
 26-Aug-1995

ACCESSIONS  
 REFERENCE  
 \*authors  
 \*journal  
 \*title

A36721  
 Webb, R.; Reddy, K.J.; Sherman, L.A.  
 J. Bacteriol. (1990) 172:5079-5088  
 Regulation and sequence of the *Synechococcus* sp. strain PCC  
 7942 drbSL operon, encoding a cyanobacterial chaperonin.

\*cross-references NCBI:90362561  
 \*accession A36721  
 \*status preliminary  
 \*molecule\_type DNA

\*residues 1-103 \*label WEB  
 \*cross-references GB:M58751; NID:q154519; PIDN:AA27313.1; PID:q154520  
 CLASSIFICATION \*superfamily chaperonin groES  
 SUMMARY \*length 103 \*molecular-weight 10742 \*checksum 6552

Query Match 35.2% Score 80; DB 2; Length 103;  
 Best Local Similarity 50.0%; Pred. No. 7.31e-01;  
 Matches 8; Conservative 7; Mismatches 1; Indels 0; Gaps 0;

DB 88 DQVLLSEKDLAVVA 103  
 QY 1 EYLLSARDVLAVVS 16

RESULT 13  
 ENTRY  
 TITLE

T11320 \*type complete  
 NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 1 -  
 Pedinorcas minor mitochondrion (SGC3)  
 \*formal\_name mitochondrion: *Pedinomonas minor*  
 16-Jul-1999 \*sequence\_revision 16-Jul-1999 \*text\_change  
 13-Aug-1995

ACCESSIONS  
 REFERENCE  
 \*authors

T11320  
 Turmel, M.; Lemieux, C.; Burger, G.; Lang, B.F.; Otis, C.;  
 Plante, I.; Gray, M.W.  
 submitted to the EMBL Data Library, December 1998  
 \*description The complete mitochondrial DNA sequences of *Nephroselmis*  
 olivacea and *Pedinomonas minor*; two radically different  
 evolutionary patterns within the green algae.

\*accession T11320  
 \*status preliminary: translated from GB/EMBL/DBJ  
 \*molecule\_type DNA  
 \*residues 1-310 \*label TJR  
 \*cross-references EMBL:AF116775; NID:q4378756; PID:q4378774;  
 PIDN:AAD19672.1

GENETICS  
 \*gene nadl  
 \*genome mitochondrion:  
 \*genetic\_code SGC3

CLASSIFICATION \*superfamily NADH dehydrogenase (ubiquinone) chain 1  
 KEYWORDS mitochondrion; NAD; oxidoreductase  
 SUMMARY \*length 310 \*molecular-weight 34938 \*checksum 2458

Query Match 35.2% Score 80; DB 2; Length 310;  
 Best Local Similarity 41.7%; Pred. No. 7.31e-01;  
 Matches 10; Conservative 8; Mismatches 8; Indels 0; Gaps 0;

DB 28 FLVLIERKVIAYIQRGGPVFGP 1  
 QY 3 YLLSARDVLAVVSKRRMKVIVAF 26

RESULT 14  
 ENTRY  
 TITLE

T11938 \*type complete  
 NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 1 -  
 Prototheca wickerhamii mitochondrion  
 \*formal\_name mitochondrion: *Prototheca wickerhamii*  
 16-Jul-1999 \*sequence\_revision 16-Jul-1999 \*text\_change  
 13-Aug-1995

ACCESSIONS  
 REFERENCE  
 \*authors  
 \*journal  
 \*title

T11938  
 Wolff, G.; Plante, I.; Lang, B.F.; Kueck, T.; Burger, G.;  
 J. Mol. Biol. (1997) 237:75-86  
 Complete sequence of the mitochondrial DNA of the chlorophyte  
 alga *Prototheca wickerhamii*. Gene content and genome  
 organization.

\*cross-references NCBI:94180393  
 \*accession T11938  
 \*status preliminary: translated from GB/EMBL/DBJ  
 \*molecule\_type DNA  
 \*residues 1-328 \*label WGL  
 \*cross-references EMBL:U02970; NID:q467843; PID:q467870;  
 PIDN:AA012657.1  
 \*experimental\_source strain 263-11

GENETICS  
 \*genome mitochondrion  
 \*note nadl

CLASSIFICATION \*superfamily NADH dehydrogenase (ubiquinone) chain 1  
 KEYWORDS mitochondrion; NAD; oxidoreductase  
 SUMMARY \*length 328 \*molecular-weight 36437 \*checksum 9332

Query Match 35.2% Score 80; DB 2; Length 328;  
 Best Local Similarity 33.3%; Pred. No. 7.31e-01;  
 Matches 10; Conservative 8; Mismatches 11; Indels 1; Gaps 1;

DB 25 FLVLAERKVLASMKRRKGNVVG 54  
 QY 3 YLLSARDVLAVVSKRR-MKVTVAPQVFGP 31

RESULT 15

ENTRY  
 TITLE

BVW478 \*type complete  
 chaperonin groES, *Mycobacterium bovis*  
 ALTERNATE\_NAMES heat shock protein, hsp70 (hsp70); immunogenic protein MP857  
 ORGANISM  
 DATE

\*formal\_name *Mycobacterium bovis*  
 31-Mar-1991 \*sequence\_revision 31-Mar-1991 \*text\_change  
 16-Jul-1999

ACCESSIONS  
 REFERENCE  
 \*authors

SC1784; A37534; C60278  
 SC1784  
 Yanaguchi, R.; Matsuo, K.; Yamazaki, A.; Nishi, S.; Teraseka,  
 K.; Yamada, T.  
 FEBS Lett. (1988) 240:115-117  
 Immunogenic protein MP857 from *Mycobacterium bovis* H39;  
 molecular cloning, nucleotide sequence and expression.

\*cross-references NCBI:89052866  
 \*accession SC1784  
 \*molecule\_type DNA  
 \*residues 1-100 \*label YAM  
 \*cross-references EMBL:X11970; NID:q44190; PIDN:CAA32149.1; PID:q4581313  
 \*accession A37534  
 \*molecule\_type protein  
 \*residues 2-21 \*label AM2

REFERENCE A60278  
 \*authors  
 \*journal

Filis, T.; Costopoulos, C.; Radford, A.J.; Bacic, A.; Wood,  
 P.R.  
 Infect. Immun. (1991) 59:800-807

#title Purification and characterization of major antigens from a  
Mycobacterium bovis culture filtrate.

#cross-references MUID:9:147217

#accession C60278

#molecule-type protein

#residues 2:15, E, 17-20 #label F1F

GENETICS

#gene

groES

#start\_codon GTG

CLASSIFICATION #superfamily chaperonin groES

KEYWORDS heat shock; molecular: chaperone; stress-induced protein

FEATURE

2-100

SUMMARY #product chaperonin groES #status predicted #label MAI

#length 100 #molecular\_weight 10950 #checksum 7378

Query Match 34.8% Score 79 DB 1: length 100:

Best Local Similarity 84.6% Pred. No. 1:0:0:0:

Matches 11: Conservative 2: Mismatches 0: Indels 0: Gaps 0:

DB 84 EYVILSARDVWG 96

1:1111111111

QY 1 EYVILSARDVLA 13

Search completed: Mon Jun 19 16:12:12 2000  
for time 1.9 secs.



PC STRAIN-H37RV;  
 RX MEDLINE: 98295987;  
 RA Cole S.L., Brosch K., Parkhill J., Garnier T., Churcher C., Harris D.,  
 RA Gordon S.V., Eigemeier K., Gas S., Barry C.F. III, Tekala F.,  
 RA Radcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
 RA Davies R., Deakin K., Fellwell T., Gentles S., Hamlin N., Holroyd S.,  
 RA Hornsby T., Jacobs K., Krogh A., McLean J., Moule S., Murphy L.,  
 RA Oliver S., Osborne J., Quail M.A., Rajandream K.A., Rogers J.,  
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares K., Sulston J.E.,  
 RA Taylor K., Whitehead S., Barrall B.G.;  
 RA "Characterizing the biology of Mycobacterium tuberculosis from the  
 RA complete genome sequence."  
 PL Nature 393:537-544(1998).  
 RN [6];  
 RP SEQUENCE OF 1-15;  
 RX MEDLINE: 92176646;  
 RA Barnes P.F., Mehra V., Riviere B., Ford S.J., Brennan P.J.,  
 RA Voegtlin M.S., Minden P., Houghton R.A., Bloom B.R., Modlin R.L.;  
 RA "Immunoreactivity of a 10-kDa antigen of Mycobacterium tuberculosis";  
 RL J. Immunol. 145:1835-1840(1992).  
 CC -- FUNCTION: BINDS TO CPN60 IN THE PRESENCE OF MG-ATP AND SUPPRESSES  
 CC THE ATPASE ACTIVITY OF THE LATTER  
 CC -- SUBUNIT: HEPTAMER OF 7 SUBUNITS ARRANGED IN A KING  
 CC (BY SIMILARITY).  
 CC -- SIMILARITY: BELONGS TO THE GROES CHAPERONIN FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 DR EMBL: X60559; CAA42908.1;  
 DR EMBL: X12598; CAA31111.1;  
 DR EMBL: M2258; AAA25343.1;  
 DR EMBL: X13739; AAA32033.1;  
 DR EMBL: Z77165; JAB01025.1;  
 DR PIR: S02727; BVNYBA;  
 DR PIR: A37166; A37166;  
 DR PIR: A47292; A47292;  
 DR PRINTS: PR00297; CHAPERONIN10;  
 DR PROSITE: PS0068; CHAPERONINS\_CPN10; 1;  
 DR PRAM: PF00186; cpn10; 1;  
 DR THERMOLIST: H3418c; 1;  
 KW Chaperone: Antigen: Heat shock;  
 FT INIT\_MET 0;  
 FT SEQUENCE 99 AA; 10669 MW; 1601285570519AFA CRC64;  
 Query Match 50.2%; Score 114; DB 1; Length 99;  
 Best Local Similarity 100.0%; Pred. No. 7.07e-06;  
 Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 DB 83 EYLLSARDVLAVSK 99  
 QY 1 EYLLSARDVLAVSK 17  
 RESULT 2  
 ID CH10\_MYCLE STANDARD; PRT: 99 AA;  
 AC P24301;  
 DT 01-MAR-1992 (Rel. 21; Created)  
 DT 01-MAR-1992 (Rel. 21; Last sequence update)  
 DT 15-DEC-1998 (Rel. 37; Last annotation update)  
 DE 10 KD CHAPERONIN (PROTEIN CPN10) (PROTEIN GROES) (10 KD ANTIGEN).  
 GN KOPB OR GROES OR CHPA OR B1620\_C3\_227 OR B229\_C3\_247;  
 OS Mycobacterium leprae;  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.  
 RN [1]  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-24.  
 RX MEDLINE: 9-1113459.

RA Mehra V.L., Bloom B.R., Bajardi A.C., Grisso A.C., Steiling P.A.,  
 RA Alland D., Convit C., Fan X., Hunter S.W., Brennan P., Rea I.B.  
 RA Modlin R.L.;  
 RA "A major T cell antigen of Mycobacterium leprae is a 10-kD heat-shock  
 RA cognate protein.";  
 RL J. Exp. Med. 175:275-284(1992).  
 RN [2];  
 RP SEQUENCE FROM N.A.;  
 RX MEDLINE: 92374850;  
 RA de Wit T.F.R., Bekelie S., Osland A., Miko T.L., Hermans P.W.M.,  
 RA van Soelingen D., Drijfhout J., Schrevelink R., Jansen A.A.M.;  
 RA Thole J.E.R.;  
 RA "Mycobacteria contain two groEL genes: the second Mycobacterium  
 RA leprae groEL gene is arranged in an operon with groES".  
 RL Mol. Microbiol. 6:1995-2007(1992).  
 RN [3];  
 RP SEQUENCE FROM N.A.;  
 RX MEDLINE: 92374850;  
 RA Smith D.R., Robison K.;  
 RL Submitted (MAR-1994) to the EMBL/GenBank/DBS databases.  
 RN [4];  
 RP X-RAY CRYSTALLOGRAPHY (3.5 ANGSTROMS).  
 RX MEDLINE: 96138402;  
 RA Wende S.C., Mehra V., Bloom B.R., Hol W.G.J.;  
 RA "Structure of the heat shock protein chaperonin-10 of Mycobacterium  
 RA leprae".  
 RL Science 271:203-207(1996).  
 RN [5];  
 RP EXTRATUM.  
 RA Wende S.C., Mehra V., Bloom B.R., Hol W.G.J.;  
 RL Science 271:1655-1655(1996).  
 CC -- FUNCTION: BINDS TO CPN60 IN THE PRESENCE OF MG-ATP AND SUPPRESSES  
 CC THE ATPASE ACTIVITY OF THE LATTER.  
 CC -- SUBUNIT: HEPTAMER OF 7 SUBUNITS ARRANGED IN A KING  
 CC (BY SIMILARITY).  
 CC -- SIMILARITY: BELONGS TO THE GROES CHAPERONIN FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 DR EMBL: X59413; NOT\_ANNOTATED\_CDS;  
 DR EMBL: Z11665; NOT\_ANNOTATED\_CDS;  
 DR EMBL: J00015; AAC4327.1;  
 DR EMBL: J00020; AAA1731.1;  
 DR PIR: JHC556; JHC556;  
 DR PIR: S25189; S25189;  
 DR PIR: LLEP; LLEP;  
 DR PRINTS: PR00247; CHAPERONIN;  
 DR PROSITE: PS0068; CHAPERONINS\_CPN10; 1;  
 DR PRAM: PF00186; cpn10; 1;  
 KW Chaperone: Antigen: Heat shock; 3D-structure.  
 FT INIT\_MET 0;  
 FT SEQUENCE 99 AA; 10669 MW; 602BBB9833F7DEB CRC64;  
 Query Match 50.2%; Score 114; DB 1; Length 99;  
 Best Local Similarity 100.0%; Pred. No. 7.07e-06;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 DB 83 EYLLSARDVLAVSK 99  
 QY 1 EYLLSARDVLAVSK 17  
 RESULT 3  
 ID CH10\_MYCAV STANDARD; PRT: 99 AA;  
 AC O86017;  
 DT 15-FEB-2000 (Rel. 39; Created)  
 DT 15-FEB-2000 (Rel. 39; Last sequence update)  
 DT 15-FEB-2000 (Rel. 39; Last annotation update)  
 DE 10 KD CHAPERONIN (PROTEIN CPN10) (PROTEIN GROES) (10 KD ANTIGEN).







CC entities and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch))

DR EMBL: D45880; RAA08298.1; 1;  
DR EMBL: U29483; AAA83440.1; 1;  
DR PIR: B39313; B39313;  
DR PRINTS: PR00297; CHAPERONIN10;  
DR PROSITE: PS00681; CHAPERONINS\_CPN10; 1;  
DR PFAM: PF00166; cpn10; 1;  
KW Chaperone.  
FT INIT\_MET 0 0  
FT CONFLICT 19 19 R -> K (IN REF. 3).  
FT CONFLICT 83 86 DGE -> APRRT (IN REF. 2).  
SV SEQUENCE 100 AA; 10865 MW; 452F3DD9CCCB0FA CRC64;

Query Match 35.2% Score 80; DB 1; Length 100;  
Best Local Similarity 73.3%; Pred. No. 1.60e-01;  
Matches 11; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

DB 85 EYVLLSERDLAVL 99  
QY 1 EYVLLSARDVLAV 15

RESULT 8  
AC CH10\_SYN6 STANDARD: PRT: 103 AA.  
DI 01-AUG-1989 (Rel. 08, Created)  
DI 01-AUG-1989 (Rel. 08, Last sequence update)  
DI 01-FEB-1996 (Rel. 33, Last annotation update)  
DI 10 KD CHAPERONIN (PROTEIN CPN10) (PROTEIN GROES).  
GN GROES.  
OS Synchococcus sp. (strain PCC 6301).  
OC Bacteria; Cyanobacteria; Chroococcales; Synchococcus.  
PN [1]  
HP SEQUENCE FROM N.A.  
RX MEDLINE: 8731173.  
RA Cozars A.L., Walker J.E.:  
RI "The organization and sequence of the genes for ATP synthase subunits  
RI in the cyanobacterium Synchococcus 6301. Support for an  
RI endosymbiotic origin of chloroplasts."  
RC J. Mol. Biol. 194:359-383(1987).  
RN [2]  
RW SIMILARITY TO CHAPERONINS.  
RX MEDLINE: 89366572.  
RA Cookson M.C., Baird P.N., Hall L.M., Coates A.R.M.:  
RI "Identification of two unknown reading frames in Synchococcus 6301  
RI as homologues of the 11k and 65k antigen genes of Mycobacterium  
RI tuberculosis and related heat shock genes in E. coli and Coxsiella  
RI burnetii."  
PL Nucleic Acids Res. 17:6392-6392(1989).  
CC -- FUNCTION: BINDS TO CPN60 IN THE PRESENCE OF MG-ATP AND SUPPRESSES  
CC THE ATPASE ACTIVITY OF THE LATTER.  
CC -- SUBUNIT: HEPTAMER OF 7 SUBUNITS ARRANGED IN A RING  
CC (BY SIMILARITY).  
CC -- SIMILARITY: BELONGS TO THE GROES CHAPERONIN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
DR EMBL: X05925; CAA29361.1; 1;  
DR PIR: S10836; BVYCGS.  
DR PRINTS: PR00297; CHAPERONIN10.  
DR PROSITE: PS00681; CHAPERONINS\_CPN10; 1;  
DR PFAM: PF00166; cpn10; 1;  
KW Chaperone.  
SV SEQUENCE 103 AA; 10811 MW; FFED1DE5F515952F CRC64;

Query Match 35.2% Score 80; DB 1; Length 103;  
Best Local Similarity 50.0%; Pred. No. 1.60e-01;  
Matches 8; Conservative 7; Mismatches 1; Indels 0; Gaps 0;

DB 88 DDYVLLSEKDLAVWA 103  
QY 1 EYVLLSARDVLAVS 16

RESULT 9  
ID CH10\_SYN7 STANDARD: PRT: 103 AA.  
AC P22880;  
DI 01-AUG-1991 (Rel. 19, Created)  
DI 01-AUG-1991 (Rel. 19, Last sequence update)  
DI 01-FEB-1996 (Rel. 33, Last annotation update)  
DI 10 KD CHAPERONIN (PROTEIN CPN10) (PROTEIN GROES).  
GN GROES.  
OS Synchococcus sp. (strain PCC 7942) (Anacystis nidulans #2).  
OC Bacteria; Cyanobacteria; Chroococcales; Synchococcus.  
PN [1]  
HP SEQUENCE FROM N.A.  
RX MEDLINE: 90368561.  
RA Webb R., Reddy K.J., Sherman L.A.:  
RI "Regulation and sequence of the Synchococcus sp. strain PCC 7942  
RI groESL operon, encoding a cyanobacterial chaperonin."  
RC J. Bacteriol. 172:5079-5080(1990).  
CC -- FUNCTION: BINDS TO CPN60 IN THE PRESENCE OF MG-ATP AND SUPPRESSES  
CC THE ATPASE ACTIVITY OF THE LATTER.  
CC -- SUBUNIT: HEPTAMER OF 7 SUB UNITS ARRANGED IN A RING  
CC (BY SIMILARITY).  
CC -- SIMILARITY: BELONGS TO THE GROES CHAPERONIN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
DR EMBL: M58751; AAA27313.1; 1;  
DR PIR: A36721; A36721.  
DR PRINTS: PR00297; CHAPERONIN10.  
DR PROSITE: PS00681; CHAPERONINS\_CPN10; 1;  
DR PFAM: PF00166; cpn10; 1;  
KW Chaperone.  
SV SEQUENCE 103 AA; 10742 MW; 89467145F515952E CRC64.

Query Match 35.2% Score 80; DB 1; Length 103;  
Best Local Similarity 50.0%; Pred. No. 1.60e-01;  
Matches 8; Conservative 7; Mismatches 1; Indels 0; Gaps 0;

DB 86 DDYVLLSEKDLAVWA 103  
QY 1 EYVLLSARDVLAVS 16

RESULT 10  
ID CH10\_MYCBO STANDARD: PRT: 99 AA.  
AC P15020;  
DI 01-APR-1990 (Rel. 14, Created)  
DI 01-AUG-1991 (Rel. 19, Last sequence update)  
DI 15-FEB-2000 (Rel. 39, Last annotation update)  
DI 10 KD CHAPERONIN (PROTEIN CPN10) (PROTEIN GROES) (IMMUNOGENIC PROTEIN  
DE MP857).  
GN MOPB OR GROES.  
OS Mycobacterium bovis.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.  
RN [1]  
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-20.  
RW STRAIN-BGG;

EX MEDLINE: 39052855.  
 KA Yamaguchi R., Matsuo K., Yamazaki A., Nagai S., Terasaka K.,  
 RA Yaraqi T.,  
 RI "Immunogenic protein MPB57 from Mycobacterium bovis BCG: molecular  
 cloning, nucleotide sequence and expression."  
 FEBS Lett. 240:115-117(1988).  
 CC -1- FUNCTION: BINDS TO CPN60 IN THE PRESENCE OF MG-ATP AND SUPPRESSES  
 THE ATPASE ACTIVITY OF THE LATTER.  
 CC -1- SUBUNIT: REMAINDER OF 7 SUBUNITS ARRANGED IN A RING  
 (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE GROS CHAPERONIN FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation.  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL: X15973; CAA32149.1;  
 EMBL: X15389; AAA25365.1;  
 PIR: S07384; BVN78.  
 PRINTS: PR002977; CHAPERONIN10.  
 PROSITE: PS00681; CHAPERONINS\_CPN10: 1.  
 PFAM: PF00166; cpn10: 1.  
 KW Chaperone; Antigen; Heat shock.  
 EC INT\_MIT  
 CC SEQUENCE 99 AA: 108-9 MW: 100552.99447AF7 CRC64.

Query Match 34.8%; Score 79; DB: Length 99;  
 Best Local Similarity 84.6%; Pred. NO: 2.35e-01;  
 Matches 11; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

CC 83 EYLLSARDVVG 95  
 CC 1 EYLLSARDVIA 13  
 CC 11  
 CC PGLYCEAST STANDARD PRT: 869 A.  
 CC P32393.  
 CC 01-OCT-1993 (Rel. 27, Created)  
 CC 01-OCT-1993 (Rel. 27, Last sequence update)  
 CC 15-FEB-2000 (Rel. 39, Last annotation update)  
 CC 1-PHOSPHATIDYL-INOSITOL-4,5-BISPHOSPHATE PHOSPHATIDYLINOSITOL-4,5-BISPHOSPHATASE  
 (EC 3.1.4.11) (GROS) THE PHOSPHATASE  
 CC PGL1 OR YPL2K6N  
 CC Saccharomyces cerevisiae (Baker's yeast).  
 CC Eukaryotic Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;  
 CC Saccharomycetaceae; Saccharomycetes.  
 CC  
 CC SEQUENCE FROM N.A.  
 CC MEDLINE: 93189586.  
 CC Yoko-O.T., Matsui Y., Yagisawa H., Nojima H., Uho I., Toh-E A.;  
 CC "The putative phosphatidyl-specific phospholipase C gene, PGL1,  
 CC of the yeast Saccharomyces cerevisiae is important for cell growth."  
 CC Proc. Natl. Acad. Sci. U.S.A. 90:1854-1858(1993).  
 CC [2]  
 CC SEQUENCE FROM N.A.  
 CC MEDLINE: 9336018.  
 CC Flick J.S., Thotter J.W.;  
 CC "Genetic and biochemical characterization of a phosphatidyl-inositol-  
 CC specific phospholipase C in Saccharomyces cerevisiae."  
 CC Mol. Cell. Biol. 13:5861-5876(1993).  
 CC [3]  
 CC SEQUENCE FROM N.A.  
 CC MEDLINE: 93309469.  
 CC Payne W.E., Fitzgerald-Hayes M.;  
 CC "A mutation in PGL1, a candidate phosphoinositide-specific  
 CC phospholipase C gene from Saccharomyces cerevisiae, causes aberrant  
 CC mitotic chromosome segregation."  
 CC Mol. Cell. Biol. 13:4351-4364(1993).

RN [4]  
 RP SEQUENCE FROM N.A.  
 RA Delius H., Hebling U.;  
 RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RA Dueterhoeft A., Floeth M., Fritz M., Hilbert H., Moestl D.  
 RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: THE PRODUCTION OF THE SECOND MESSENGER MOLECULES  
 DIACYLGLYCEROL (DAG) AND INOSITOL 1,4,5-TRISPHOSPHATE (IP3) IS  
 MEDIATED BY ACTIVATED PHOSPHATIDYLINOSITOL-SPECIFIC PHOSPHATASE  
 C ENZYMES. THIS ENZYME IS ALSO REQUIRED FOR CELL GROWTH.  
 CC -1- CATALYTIC ACTIVITY: 1-PHOSPHATIDYL-D-MYO-INOSITOL 4,5-BISPHOSPHATE  
 + H2O = D-MYO-INOSITOL 1,4,5-TRISPHOSPHATE + DIACYLGLYCEROL  
 CC -1- SIMILARITY: DOMAINS X AND Y ARE CONSERVED IN DIFFERENT FORMS OF  
 PLC AND ARE ESSENTIAL FOR CATALYTIC ACTIVITY.  
 CC -1- SIMILARITY: CONTAINS 1 C2 DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 1 EF-HAND CALCIUM-BINDING DOMAIN.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation.  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL: D12738; BAA22230.1;  
 EMBL: L13036; AAA99927.1;  
 EMBL: S63468; CAB32216.1;  
 EMBL: Z73224; CAA98004.1;  
 EMBL: Z73223; CAA98003.1;  
 PIR: A47257; A47257.  
 PIR: A48142; A48142.  
 DR HSP: P10688; ID32.  
 DR SGD: L000148; PCL  
 DR PRINTS: PR00390; PHPLIPASEC.  
 DR PROSITE: PS00018; EF\_HAND: 1.  
 DR PROSITE: PS50004; C2\_DOMAIN: 2.  
 DR PROSITE: PS50007; PIP2\_X\_DOMAIN: 1.  
 DR PROSITE: PS50008; PIP2\_Y\_DOMAIN: 1.  
 DR PFAM: PF00036; ehand: 1.  
 DR PFAM: PF00168; C2: 1.  
 DR PFAM: PF00387; PI-PLC-Y: 1.  
 DR PFAM: PF03388; PI-PLC-X: 1.  
 KW Hydrolyase; Lipid binding; Phosphatase; Phospholipase; Phosphatidyl-  
 CC CALCIUM  
 CC -1- FUNCTION: THIS ENZYME IS REQUIRED FOR CELL GROWTH.  
 CC -1- SIMILARITY: CONTAINS 1 C2 DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 1 EF-HAND CALCIUM-BINDING DOMAIN.  
 CC -1- ACT SITE  
 CC ACT SITE 734 846  
 CC BY SIMILARITY  
 CC BY SIMILARITY  
 CC BY SIMILARITY  
 CC BY SIMILARITY  
 CC CONFLICT 139 159  
 CC CONFLICT 181 181  
 CC CONFLICT 581 581  
 CC CONFLICT A -> V (IN REF. 2).  
 CC SEQUENCE 869 AA: 100547 MW: 1386751259405590 CRC64.

Query Match 34.8%; Score 79; DB: Length 869;  
 Best Local Similarity 29.6%; Pred. NO: 2.35e-01;  
 Matches 8; Conservative 12; Mismatches 5; Indels 2; Gaps 2;

DB 702 Y-VLKPKKLPVVTAKM-IPLIYERF 726  
 QY 3 Y-LILSARDVLAVSKRRKRVIVAFNQF 29  
 ID CHIC\_SYNVU STANDARD: PRT: 102 AA.  
 AC Q50322;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE 10 KS CHAPERONIN (PROTEIN CPN10) (PROTEIN GROS).

GN GROES  
OS Synecococcus vulcanus.  
CC Bacteria: Cyanobacteria, Chroococcales: Synecococcus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 98094212.  
RA Tanaka N., Miyama T., Nakamoto H.:  
RT "Cloning, characterization and functional analysis of groESL operon  
from thermophilic cyanobacterium Synecococcus vulcanus".  
RL Biochim. Biophys. Acta 1343:335-348(1997).  
CC -!- FUNCTION: BINDS TO CPN60 IN THE PRESENCE OF MG-ATP AND SUPPRESSES  
THE ATPASE ACTIVITY OF THE LATTER.  
CC -!- SUBUNIT: HEPTAMER OF 7 SUBUNITS ARRANGED IN A RING  
(BY SIMILARITY).  
CC -!- SIMILARITY: BELONGS TO THE GROES CHAPERONIN FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: D78139; BAA23816.1.  
DR PRINTS: PRO0297: CHAPERONIN10.  
DR PROSITE: PS00681: CHAPERONINS\_CPN10; 1.  
DR PFAM: PF00156; cpn10; 1.  
KW Chaperone.  
FT INIILMET 0 0 BY SIMILARITY.  
SQ SEQUENCE 102 AA: 10783 MW: 87CC3517298BFF67 CRC64:  
  
Query Match 34.4% Score 78: DB 1: Length 102:  
Best Local Similarity 50.0% Pred. No. 3.44e-01:  
Matches 8: Conservative 7: Mismatches 1: Indels 0: Gaps 0:  
  
Db 87 EDVILSEKDIILAIIG 102  
QY 1 EYVILSARDVLAVS 16  
IIIIII:IIIIII:  
-----  
RESULT 13  
ID CH10\_CVAPA STANDARD: PRI: 103 AA.  
AC Q3776:  
DT 01-NOV-1997 (Rel. 35, Created)  
DI 01-NOV-1997 (Rel. 35, Last sequence update)  
DE 10 KD CHAPERONIN (PROTEIN CPN10) (PROTEIN GROES).  
OS Cyatophora paradoxa.  
CC Cyanelle.  
CC Eukaryota: Glaucocystophyceae; Cyatophora-aeo: Cyatophora.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=LBS35 / PRINGSHEIM;  
RA Stirewalt V.L., Michalowski C.B., Luffelhardt W., Bonnett H.J.,  
RL Bryant D.A.:  
RT Submitted (OCT-1995) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: BINDS TO CPN60 IN THE PRESENCE OF MG-ATP AND SUPPRESSES  
THE ATPASE ACTIVITY OF THE LATTER (BY SIMILARITY).  
CC -!- SUBUNIT: HEPTAMER OF 7 SUBUNITS ARRANGED IN A RING  
(BY SIMILARITY).  
CC -!- SIMILARITY: BELONGS TO THE GROES CHAPERONIN FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: J30821; AAA81173.1.  
DR

DR EMBL: U30821; AAA81293.1.  
DR PRINTS: PRO0297: CHAPERONIN10.  
DR PROSITE: PS00681: CHAPERONINS\_CPN10; 1.  
DR PFAM: PF00166; cpn10; 1.  
KW Chaperone; Cyanelle.  
SQ SEQUENCE 103 AA: 11240 MW: FF768479C0A90520 CRC64:  
  
Query Match 34.4% Score 78: DB 1: Length 103:  
Best Local Similarity 50.0% Pred. No. 3.44e-01:  
Matches 8: Conservative 7: Mismatches 1: Indels 0: Gaps 0:  
  
Db 89 EYVILSEKDIILAIIZ 103  
QY 1 EYVILSARDVLAVS 16  
IIIIII:IIIIII:  
-----  
RESULT 14  
ID NTPOLENTHR STANDARD: PRI: 230 AA.  
AC P43435:  
DT 01-NOV-1995 (Rel. 32, Created)  
DI 01-NOV-1995 (Rel. 32, Last sequence update)  
DE V-TYPE SODIUM ATP SYNTHASE SUBUNIT D (EC 3.6.1.34) (NA(+)).  
OS TRANSLOCATING ATPASE SUBUNIT D.  
CC NTPO.  
CC Enterococcus hirae.  
CC Bacteria: Firmicutes: Bacillus/Clostridium group: Enterococcaceae:  
CC Enterococcus.  
CC [1]  
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-14.  
RX STRAIN=ATCC 9790;  
RX MEDLINE: 94209269.  
RA Takase K., Kakinuma S., Yamato I., Konishi K., Igarashi K.:  
RT Sequencing and characterization of the atp gene cluster for  
vacuolar-type Na(+)-translocating ATPase of Enterococcus hirae.  
RL J. Biol. Chem. 269:11037-11044(1994).  
CC -!- FUNCTION: INVOLVED IN ATP-DRIVEN SODIUM EXTRUSION.  
CC -!- SIMILARITY: BELONGS TO A FAMILY THAT GROUPS V-ATPASE SUBUNIT D.  
CC V-TYPE SODIUM ATPASE NTPO AND ARCHEAL ATPASE SUBUNIT D.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: D17472; BAA34275.1.  
CC Hydrolase: Hydrogen ion transport  
KW  
SQ SEQUENCE 230 AA: 27293 MW: 950021364E58F0: CRC64:  
  
Query Match 34.4% Score 78: DB 1: Length 230:  
Best Local Similarity 40.0% Pred. No. 3.44e-01:  
Matches 12: Conservative 7: Mismatches 10: Indels 1: Gaps 1:  
  
Db 78 DELLALPAENVYLSVVEKNIMSVKVLMMNF 107  
QY 1 EYVILSARDV-LAVVSKRRMKVIVAFNOF 29  
IIIIII:IIIIII:IIIIII:  
-----  
RESULT 15  
ID CH10\_LEPIN STANDARD: PRI: 96 AA.  
AC P35472:  
DT 01-JUN-1994 (Rel. 29, Created)  
DI 01-JUN-1994 (Rel. 29, Last sequence update)  
DE 15-DEC-1998 (Rel. 37, Last annotation update)  
DE 10 KD CHAPERONIN (PROTEIN CPN10) (PROTEIN GROES) (HEAT SHOCK 10 KD  
DE PROTEIN).  
GN GROES OR HSP10.  
OS Leptospira interrogans.  
CC Bacteria: Spirochaetales: Leptospiraceae: Leptospira.

```

(1)
SEQUENCE FROM N.A.
STRAIN-SEROVAR COPENHAGEN / WIJNBORG:
MEDLINE: 9313752.
Ballard S.A., Seders R.P., Breutink-Piym N.M., Eylar J.A.M.,
Faine S., Adler B.:
"molecular analysis of the hsp (groE) operon of Leptospira
interrogans serovar Copenhagen";
Mol. Microbiol. 8:739-751(1993).
(2)
SEQUENCE FROM N.A.
STRAIN-SEROVAR LAI / HN-1:
Kim M.J., Ahn S.Y.:
Submitted (Nov-1997) to the EMBL/Genbank/DBJ databases.
-1- FUNCTION: BINDS TO CPN50 IN THE PRESENCE OF MG-ATP AND SUPPRESSES
THE ATPASE ACTIVITY OF THE LATTER.
-1- SUBUNIT: HEPTAMER OF 7 SUBUNITS ARRANGED IN A RING
BY SIMILARITY.
-1- INDUCTION: BY HEAT SHOCK.
-1- SIMILARITY: BELONGS TO THE GROES CHAPERONIN FAMILY.
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
EMBL: U14682; AAA71991.1;
EMBL: AF022910; AAR86964.1;
PIR: S34937; S34937.
PRINTS: PR00097; CHAPERONIN10.
PROSITE: PS01681; CHAPERONINS_CPN10; 1.
PFAM: PF00166; cpn10; 1.
Chaperone: Heat Shock.
SEQUENCE 96 AA: 10562 MW: 528508079D8F3467 CRC64:
Query Match 33.5% Score 76; DB 1; Length 96;
Best Local Similarity 62.5%; Pred. No. 7326-01;
Matches 10; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
LE 8: EYDIIRESQILAVVKK 96
10: 10: 10: 10: 10: 10: 10: 10:
DY 2 EYDIISSAPGVAVVSK 17
Search completed: Mon Jun 19 16:11:06 2000
Tot time : 17 secs.

```

\*\*\*\*\*  
WATERFALL (TM)  
\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit,  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MParch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Mon Jun 19 16:11:24 2000: MasPar time 9.82 seconds  
Tabular output not generated. 225,834 Million cell updates/sec

Title: >US-09-142-524A-10  
Description: (1-32) from US09142524A.pep  
Perfect Score: 227  
Sequence: 1 EYLILSARDVLAVVSKRMKV-VAFNQEGPN 32

Scoring table: PAM 150  
Gap 11

Searched: 225878 seqs, 59334122 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: streneb112  
1:sp\_archaea 2:sp\_bacteria 3:sp\_fungi 4:sp\_human  
5:sp\_invertebrate 6:sp\_mammal 7:sp\_mhc 8:sp\_organelle  
9:sp\_phase 10:sp\_plant 11:sp\_rodent 12:sp\_unclassified  
13:sp\_vertebrate 14:sp\_virus

Statistics: Mean 33.599; Variance 54.501; scale 0.616

pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description            | Pred. No. |
|------------|-------|-------------|--------|-------|------------------------|-----------|
| 1          | 114   | 50.2        | 100    | 2     | 10 KD CHAPERONIN (PROT | 2.33e-07  |
| 2          | 98    | 43.2        | 357    | 10    | POLLEN MAJOR ALLERGEN  | 2.42e-04  |
| 3          | 98    | 43.2        | 357    | 10    | CHAOI PRECURSOR.       | 2.42e-04  |
| 4          | 97    | 42.7        | 459    | 10    | PUTATIVE PECTINASE     | 3.68e-04  |
| 5          | 92    | 40.5        | 354    | 10    | PECTINASE-LIKE PRO     | 2.90e-03  |
| 6          | 91    | 40.1        | 354    | 10    | PECTINASE-LIKE PRO     | 4.35e-03  |
| 7          | 88    | 38.8        | 242    | 5     | ERCC1-LIKE EXCISION RE | 1.45e-02  |
| 8          | 88    | 38.8        | 390    | 10    | F12F1.22 PROTEIN       | 1.45e-02  |
| 9          | 84    | 37.0        | 455    | 10    | PUTATIVE PECTINASE     | 7.02e-02  |
| 10         | 80    | 35.2        | 310    | 8     | NADH DEHYDROGENASE SUB | 3.25e-01  |
| 11         | 80    | 35.2        | 398    | 8     | NADH DEHYDROGENASE (UB | 3.25e-01  |
| 12         | 79    | 34.8        | 398    | 10    | PECTINASE-LIKE PRO     | 4.73e-01  |
| 13         | 78    | 34.4        | 330    | 8     | NADH-UBIQUINONE OXIDOR | 6.88e-01  |
| 14         | 77    | 33.9        | 305    | 2     | PUTATIVE OXIDOREDUCTAS | 9.96e-01  |
| 15         | 77    | 33.9        | 768    | 5     | F55H12.1 PROTEIN       | 9.96e-01  |
| 16         | 76    | 33.5        | 103    | 2     | 10 KD CHAPERONIN (PROT | 1.44e-00  |
| 17         | 76    | 33.5        | 323    | 10    | MITOCHONDRIAL NADH DEH | 1.44e-00  |
| 18         | 75    | 33.5        | 325    | 8     | NADH DEHYDROGENASE SUB | 1.44e-00  |
| 19         | 75    | 33.5        | 331    | 8     | NADH DEHYDROGENASE (UB | 1.44e-00  |
| 20         | 75    | 33.0        | 96     | 2     | HYPOTHETICAL 11.4 KD P | 2.07e-00  |

|    |    |      |      |    |        |                         |          |
|----|----|------|------|----|--------|-------------------------|----------|
| 21 | 75 | 33.0 | 225  | 1  | 098715 | 225AA LONG HYPOTHETICAL | 2.37e-00 |
| 22 | 75 | 33.0 | 4513 | 2  | 092G15 | TYPE I POLYKETIDE SYNTH | 2.37e-00 |
| 23 | 74 | 32.6 | 302  | 5  | 017137 | F31F4.3 PROTEIN         | 2.97e-00 |
| 24 | 74 | 32.6 | 1006 | 4  | 015197 | EPH-FAMILY RECEPTOR PR  | 2.97e-00 |
| 25 | 73 | 32.2 | 88   | 2  | 05C304 | GROES (FRAGMENT)        | 4.24e-00 |
| 26 | 73 | 32.2 | 129  | 8  | 095867 | NADH-UBIQUINONE OXIDOR  | 4.24e-00 |
| 27 | 73 | 32.2 | 226  | 10 | 011667 | PUTATIVE PECTINASE      | 4.24e-00 |
| 28 | 73 | 32.2 | 401  | 10 | 014557 | PECTINASE-LIKE PRO      | 4.24e-00 |
| 29 | 73 | 32.2 | 517  | 10 | 086729 | PUTATIVE CYTOCHROME P4  | 4.24e-00 |
| 30 | 73 | 32.2 | 733  | 5  | 022561 | 119H10.8 PROTEIN        | 4.24e-00 |
| 31 | 73 | 32.2 | 841  | 2  | 095709 | RELA PROTEIN            | 4.24e-00 |
| 32 | 73 | 32.2 | 2823 | 5  | 045614 | 122A3.6 PROTEIN (FRAGM  | 4.24e-00 |
| 33 | 72 | 31.7 | 301  | 2  | 053198 | OXIDOREDUCTASE          | 6.05e-00 |
| 34 | 72 | 31.7 | 374  | 2  | 012373 | CAP6G                   | 6.05e-00 |
| 35 | 72 | 31.7 | 374  | 2  | 012373 | CAP6G                   | 6.05e-00 |
| 36 | 72 | 31.7 | 438  | 10 | 043862 | POLLEN SPECIFIC PECTIN  | 6.05e-00 |
| 37 | 72 | 31.7 | 1014 | 11 | 008644 | EPH/EK RECEPTOR LIKE    | 6.05e-00 |
| 38 | 71 | 31.3 | 369  | 13 | 091790 | DUAL SPECIFICITY PROTE  | 6.59e-00 |
| 39 | 71 | 31.3 | 405  | 10 | 024416 | PECTINASE-LIKE PRO      | 6.59e-00 |
| 40 | 71 | 31.3 | 450  | 10 | 040319 | PECTINASE-LIKE PRO      | 6.59e-00 |
| 41 | 71 | 31.3 | 465  | 4  | 094889 | KIAA0795 PROTEIN (FRAG  | 6.59e-00 |
| 42 | 70 | 30.8 | 482  | 5  | 022329 | COSM10 TJH8             | 1.22e-01 |
| 43 | 70 | 30.8 | 506  | 14 | 011997 | GLYCOPROTEIN PRECURSOR  | 1.22e-01 |
| 44 | 70 | 30.8 | 509  | 14 | 011999 | GLYCOPROTEIN PRECURSOR  | 1.22e-01 |
| 45 | 70 | 30.8 | 1106 | 13 | 079726 | RECEPTOR TYROSINE KINA  | 1.22e-01 |

ALIGNMENTS

RESULT 1  
ID 086017 PRELIMINARY: PRI: 100 AA.  
AC 086017:  
DT 01-NOV-1998 (TRENBLREL, 08, Created)  
DI 01-NOV-1998 (TRENBLREL, 08, Last sequence update)  
DI 01-NOV-1998 (TRENBLREL, 12, Last annotation update)  
DE 10 KD CHAPERONIN (PROTEIN C.N.0) (PROTEIN GROES).  
GN GROES.  
OS Mycobacterium avium, and Mycobacterium paratuberculosis.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Corynebacteriaceae; Mycobacteriaceae.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-485 TYPE 21.  
RA CRET, R., PIETROBONO R., F., KIN, L., OREFICI G.  
RT "Sequence of the GROES1 protein of Mycobacterium avium comprising the  
RT gene encoding the cpn10 protein and a portion of the gene encoding the  
RT cpn60-1 protein."  
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC SPECIES-EM PARATUBERCULOSIS STRAIN-ATCC 19698, ATCC 25241  
RA GEBB A.J., PRUTHINGHAM R.  
RT "The GroES antigens of Mycobacterium avium and Mycobacterium  
RT paratuberculosis."  
RL CC  
CC FUNCTION: BINDS TO CPN60 IN THE PRESENCE OF MG-ATP AND SUPPRESS  
CC THE ATPASE ACTIVITY OF THE LATTER (BY SIMILARITY).  
CC -- SUBUNIT: HEPTAMER OF 7 SUBUNITS ARRANGED IN A RING (BY  
CC SIMILARITY).  
CC -- SIMILARITY: BELONGS TO THE GROES CHAPERONIN FAMILY.  
DR EMBL: AF079544; AAC3192.1;  
DR EMBL: AF071839; AAD23277.1;  
DR EMBL: AF071828; AAD23276.1;  
DR PROSITE: PS00681; CHAPERONINS\_CPN10; 1.  
DR PFAM: PF00166; cpn10; 1.  
DR PRINTS: PR00297; CHAPERON-N10.  
KW Chaperone; Heat shock.  
SQ SEQUENCE 100 AA; 10748 MW; 0C4954DA CRC32;

Query Match 50.2%; Score 114; DB 2; Length 100;  
Best Local Similarity 100.0%; Pred. No. 2.33e-07;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Caps 0;

Db 84 EYLLISARDVLAVWSK 100  
 QY 1 EYLLISARDVLAVWSK 17

RESULT 2 PRELIMINARY: PRT: 367 AA.  
 ID Q92N7: Q92N7: 01-FEB-1997 (TEMBLrel. 02, Created);  
 AC Q92N7: 01-FEB-1997 (TEMBLrel. 02, Last sequence update);  
 DT 01-MAY-1999 (TEMBLrel. 10, Last sequence update);  
 DT 01-MAY-1999 (TEMBLrel. 10, Last sequence update);  
 DE POLLEN MAJOR ALLEGEN 1-2;  
 OS Juniperus ashei (Ozark white cedar);  
 OC Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta;  
 OC e-phylllophytes: Spermatophyta: Coniferopsida: Coniferales;  
 OC Taxodiaceae: Juniperus.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MIDOR-HORIUT: I.M., GOLDBLUM R.M., KUKOSKY A., WOOD T.G.,  
 RA BROOKS E.G.;  
 RT "Molecular cloning of mountain cedar (Juniperus ashei) pollen major  
 RT allergen, Jun a 1";  
 RT Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: A296563; RAD03609.1;  
 DR EMBL: AF06562; RAD03608.1;  
 DR EMBL: AF06562; RAD03608.1;  
 DR MENDEL: 36544; Junas:1088;36544;  
 DR MENDEL: 36544; Junas:1088;36544;  
 DR PRINTS: PR00807; AMBALLERG.  
 KW Lysase.  
 SQ SEQUENCE 367 AA: 39824 MW: 4C2DB630 CRC32:

Query Match 43.2% Score 98: DB 10: Length 367;  
 Best Local Similarity 86.7% Pred. No. 2,42e-04;  
 Matches 13: Conservative 1: Mismatches 1: Indels 0: Gaps 0;

Db 232 KSMKVTVAFNQFGPN 246  
 QY 18 RRMKVTVAFNQFGPN 32

RESULT 3 PRELIMINARY: PRT: 376 AA.  
 ID Q56385: Q56385: 01-FEB-1997 (TEMBLrel. 02, Created);  
 AC Q56385: 01-FEB-1997 (TEMBLrel. 02, Last sequence update);  
 DT 01-FEB-1997 (TEMBLrel. 02, Last sequence update);  
 DT 01-NOV-1999 (TEMBLrel. 12, Last annotation update);  
 DE CHAOI FRECURSOR.  
 OS Chamaecyparis obtusa.  
 OC Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta;  
 OC e-phylllophytes: Spermatophyta: Coniferopsida: Coniferales;  
 OC Taxodiaceae: Chamaecyparis.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA SUZUKI M., KOMIYAMA N., ITOH M., ITOH H., SONE T., KUNO K., TAKAGI I.,  
 RA OHTA N.;  
 RT "Purification, characterization and molecular cloning of Chamaecyparis  
 RT major allergen of Chamaecyparis obtusa (Japanese cypress) pollen";  
 RT Mol. Immunol. 33:451-460(1996).  
 DR EMBL: D45404; BAA08246.1;  
 DR MENDEL: 7626; Chaoi:1088;7626.  
 DR PFAM: PF00544; pec\_lyase.1  
 DR PRINTS: PR00807; AMBALLERG.  
 KW Signal.  
 FT SIGNAL  
 FT CHAIN  
 SQ SEQUENCE 375 AA: 40258 MW: A098452 CRC32:

Query Match 43.2% Score 98: DB 10: Length 375;  
 Best Local Similarity 86.7% Pred. No. 2,42e-04;  
 Matches 13: Conservative 1: Mismatches 1: Indels 0: Gaps 0;

Db 232 KSMKVTVAFNQFGPN 246  
 QY 18 RRMKVTVAFNQFGPN 32

QY 18 RRMKVTVAFNQFGPN 32

RESULT 4 PRELIMINARY: PRT: 459 AA.  
 ID Q23665: Q23665: 01-JAN-1998 (TEMBLrel. 05, Created);  
 AC Q23665: 01-JAN-1998 (TEMBLrel. 05, Last sequence update);  
 DT 01-JAN-1998 (TEMBLrel. 05, Last sequence update);  
 DT 01-NOV-1999 (TEMBLrel. 12, Last annotation update);  
 DE PECTATE PECTATE LYASE.  
 OS Arabidopsis thaliana (Mouse ear cress);  
 OC Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta;  
 OC e-phylllophytes: Spermatophyta: Magnoliophyta: eudicotyledons;  
 OC core eudicots: Rosidae: euro: 11: Brassicales: Brassicaceae;  
 OC Arabidopsis.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE: 97422403.  
 RA KULIKAUSSAS R., MCCORMICK S.;  
 RT "Identification of the tobacco and Arabidopsis homologues of the  
 RT pollen-expressed LAI59 gene of tomato";  
 RT Plant Mol. Biol. 34:809-814(1997).  
 DR EMBL: U83613; AAB97959.1;  
 DR MENDEL: 25559; Arabid:1088;25559;  
 DR PFAM: PF00544; pec\_lyase.1;  
 DR PRINTS: PR00807; AMBALLERG.  
 KW Lysase.  
 SQ SEQUENCE 459 AA: 51420 MW: 41E590B CRC32:

Query Match 42.7% Score 97: DB 10: Length 459;  
 Best Local Similarity 46.7% Pred. No. 3,58e-04;  
 Matches 14: Conservative 9: Mismatches 5: Indels 2: Gaps 2;

Db 298 QEVMLFGARCE-HVIDKK-MQITVAFNHES 325  
 QY 1 EYLLISARDVLAVWSKRRKVTVAFNHES 33

RESULT 5 PRELIMINARY: PRT: 394 AA.  
 ID Q65457: Q65457: 01-AUG-1998 (TEMBLrel. 07, Created);  
 AC Q65457: 01-AUG-1998 (TEMBLrel. 07, Last sequence update);  
 DT 01-AUG-1998 (TEMBLrel. 07, Last sequence update);  
 DT 01-NOV-1999 (TEMBLrel. 12, Last annotation update);  
 DE PECTATE LYASE LIKE PROTEIN.  
 OS Arabidopsis thaliana (Mouse ear cress);  
 OC Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta;  
 OC e-phylllophytes: Spermatophyta: Magnoliophyta: eudicotyledons;  
 OC core eudicots: Rosidae: EUSTOIDS 11: Brassicales: Brassicaceae;  
 OC Arabidopsis.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA BEVAN M., WEDLER H., WAMBUTT R., BANCROFT I., MEWES H.W., MAYER K.,  
 RA SCHUELLER C.;  
 RT Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.  
 OS Arabidopsis thaliana (Mouse ear cress);  
 OC Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta;  
 OC e-phylllophytes: Spermatophyta: Magnoliophyta: eudicotyledons;  
 OC core eudicots: Rosidae: EUSTOIDS 11: Brassicales: Brassicaceae;  
 OC Arabidopsis.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA EU ARABIDOPSIS SEQUENCING PROJECT;  
 RA Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AL022140; CAA18112.1;  
 DR MENDEL: 29068; Arabid:1088;29068.  
 DR PFAM: PF00544; pec\_lyase.1;  
 DR PRINTS: PR00807; AMBALLERG.  
 KW Lysase.  
 SQ SEQUENCE 394 AA: 43299 MW: 8B9A93C9 CRC32:

Query Match 40.5% Score 92: DB 10: Length 394;  
 Best Local Similarity 70.6% Pred. No. 2,90e-03;  
 Matches 12: Conservative 2: Mismatches 3: Indels 0: Gaps 0;

Db 251 VKDVKMKVTVAFNHFGP 267



|||||:|||||

QY 15 VSKRMKVTVAFNFGSP 31

RESULT 6  
ID O65456 PRELIMINARY: PRT: 394 AA.  
AC O65456  
DT 01-AUG-1998 (TREMblrel. 07, Created)  
DT 01-AUG-1998 (TREMblrel. 07, Last sequence update)  
DE F12F1.22 PROTEIN.  
DE F12F1.22  
GN F12F1.22  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC eukaryotes; Spermatophyta; Magnoliophyta; eudicotyledons;  
OC core eudicots; Rosidae; eustosids II; Brassicales; Brassicaceae;  
OC Arabidopsis.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA BEVAN M., WEDLER H., WAMBOTT R., BANCROFT I., MEWES H.W., MAYER K.,  
RA SCHUELLER C.  
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA ED ARABIDOPSIS SEQUENCING PROJECT;  
RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL: A022140; CAAL9111.1;  
DR MENDEL: 29747; Arabid1088;29747.  
DR PFAM: PF00544; pec\_lyase; 1.  
DR PRINTS: PRO0807; AMBALLERGEN.  
KW Lyase.  
SQ SEQUENCE 394 AA: 43476 MW: 95399178 CRC32:

Query Match 40.1% Score 91: DB 10: Length 394:  
Best Local Similarity 84.6% Pred. No. 4.35e-03:  
Matches 11: Conservative 2: Mismatches 0: Indels 0: Gaps 0:

DB 255 KMKVTVAFNFGSP 267  
QY 19 KMKVTVAFNFGSP 31  
|||||:|||||

RESULT 7  
ID O56136 PRELIMINARY: PRT: 242 AA.  
AC O56136  
DT 01-MAY-1999 (TREMblrel. 10, Created)  
DT 01-MAY-1999 (TREMblrel. 10, Last sequence update)  
DE ERCC1-LIKE EXCISION REPAIR PROTEIN.  
GN PFG0160W.  
OS Plasmodium falciparum.  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA GARDNER M.J., TETTELIN H., CARUCCI D.J., CUMMINGS L.M., ARAVIND L.,  
RA KOONIN E.V., SHALLOM S., MASON T., YU K., FUJII C., PEPERSON J.,  
RA SHEN K., JING J., ASTON C., LAI Z., SCHWARTZ D.C., PERTEA M.,  
RA SALZBERG S., ZHOU S., SUTON G.G., CLAYTON R., WHITE C., SMITH H.O.,  
RA FRASER C.M., ADAMS M.D., VENTER J.C., HOFFMAN S.L.;  
RT \*Chromosome 2 sequence of the human malaria parasite Plasmodium  
falciparum.  
RL Science 282:1126-1132(1998).  
DR EMBL: AE001477; AAC71822.1;  
SQ SEQUENCE 242 AA: 28287 MW: A590AD3C CRC32:

Query Match 38.8% Score 88: DB 5: Length 242:  
Best Local Similarity 40.6% Pred. No. 1.45e-02:  
Matches 13: Conservative 8: Mismatches 11: Indels 0: Gaps 0:

DB 45 EYLLSARDVLAIVVSKRMKVTVAFNFGSP 76  
QY 1 EYLLSARDVLAIVVSKRMKVTVAFNFGSP 32  
|||||:|||||

RESULT 8  
ID O65388 PRELIMINARY: PRT: 390 AA.  
AC O65388  
DT 01-AUG-1998 (TREMblrel. 07, Created)  
DT 01-AUG-1998 (TREMblrel. 07, Last sequence update)  
DE F12F1.22 PROTEIN.  
DE F12F1.22  
GN F12F1.22  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC eukaryotes; Spermatophyta; Magnoliophyta; eudicotyledons;  
OC core eudicots; Rosidae; eustosids II; Brassicales; Brassicaceae;  
OC Arabidopsis.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA VISITSKAIA V.S., OSBORNE B.L., SCHWARTZ J.R., TORIEMI M., YU G.,  
RA KWAN A., OUI O., LIU S., BUEHLER E., CONWAY A.B., CONWAY A.R.,  
RA DEWAR K., FENG S., KIM C., KURTZ D., PALM C.J., LI Y., SHINN P.,  
RA SUN H., DAVIS R.W., ECKER J.R., FEDERSPIEL N.A., THEODOSIS A.;  
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AC002131; AAC17625.1;  
DR MENDEL: 29024; Arabid1088;29024.  
DR PFAM: PF00544; pec\_lyase; 1.  
DR PRINTS: PRO0807; AMBALLERGEN.  
SQ SEQUENCE 390 AA: 43354 MW: E0F2FB4 CRC32:

Query Match 38.8% Score 88: DB 10: Length 390:  
Best Local Similarity 58.8% Pred. No. 1.45e-02:  
Matches 10: Conservative 4: Mismatches 3: Indels 0: Gaps 0:

DB 241 VEDKIMRVIAFNFGSP 257  
QY 15 VSKRMKVTVAFNFGSP 31  
|||||:|||||

RESULT 9  
ID O64510 PRELIMINARY: PRT: 455 AA.  
AC O64510  
DT 01-AUG-1998 (TREMblrel. 07, Created)  
DT 01-AUG-1998 (TREMblrel. 07, Last sequence update)  
DE PUTATIVE PERIATE LYASE.  
GN L28F5.14  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC eukaryotes; Spermatophyta; Magnoliophyta; eudicotyledons;  
OC core eudicots; Rosidae; eustosids II; Brassicales; Brassicaceae;  
OC Arabidopsis.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA ROUSLEY S.D., LIN X., KETCHUM K.A., CROSBY M.L., BRANDON R.C.,  
RA SYKES S.M., KAUL S., MASON T.M., KERLAVAGE A.R., ADAMS M.D.,  
RA SOMERVILLE C.R., VENTER J.C.;  
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AC002521; AAC0350.1;  
DR MENDEL: 28505; Arabid1088;28505.  
DR PFAM: PF00544; pec\_lyase; 1.  
DR PRINTS: PRO0807; AMBALLERGEN.  
KW Lyase.  
SQ SEQUENCE 455 AA: 51257 MW: EBA0B82D CRC32:

Query Match 37.0% Score 84: DB 10: Length 455:  
Best Local Similarity 36.7% Pred. No. 7.02e-02:  
Matches 11: Conservative 11: Mismatches 6: Indels 2: Gaps 2:

DB 294 DEVMLFGGTNK-DVIDKK-MQITVAFNFGSP 321  
QY 1 EYLLSARDVLAIVVSKRMKVTVAFNFGSP 30  
|||||:|||||



```

RESULT 14
ID Q9X7I6 PRELIMINARY: PRT: 305 AA.
AC Q9X7I6
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DI 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DE 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE PUTATIVE OXIDOREDUCTASE.
GN SC5H1.28C.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2)
RA OLIVER K., HARRIS D.
RT "A set of ordered cosmids and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome."
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2)
RA JAMES K.D., PARKHILL J., BARRELL B.G., PAJANDREAM M.A.
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2)
RA MEDLINE: 9700035.
RA REDENBACH M., KIESER H.M., DENAPATE D., EICHNER A., CULLUM J.,
RA KINASHI H., HOPWOOD D.A.
RT "A set of ordered cosmids and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome."
RL NCI Microbiol. 21:77-95(1996).
OR EMBL: AL049863; CAB42953.1;
SC SEQUENCE 305 AA: 32531 MW: 5445A89 CRC32:

```

```

Query Match 33.9% Score 77: DB 2: Length 305;
Best Local Similarity 41.9% Pred. No. 9,96e-01;
Matches 13: Conservative 5: Mismatches 13: Indels 0: Gaps 0:
ID 127 EENLRQGRDHLDVYVYLRMRQDSVAEFGA 157
      1 2 3 4 5 6 7 8 9 10 11 12 13
QY 1 REYLLSARDVLAIVYVKRMRKVIVAFNFGP 31

```

```

RESULT 15
ID P98890 PRELIMINARY: PRT: 768 AA.
AC P98890
DT 01-MAY-1997 (TrEMBLrel. 08, Created)
DI 01-MAY-1997 (TrEMBLrel. 08, Last sequence update)
DE 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE F58H12.1 PROTEIN.
GN F58H12.1
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematozoa; Secernentea; Rhabditia; Rhabditidae;
OC Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN [1]
RP SEQUENCE FROM N.A.
RA DOBSON R.
RC Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC MEDLINE: 94150718.
RA WILSON R., AINSOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIEKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.

```

```

RT elegans."
RL Nature 368:32-38(1994).
DR EMBL: Z81091; CAB03142.1;
DR PFAM: PF00209; SNF: 2
DR PRINTS: PRO0176; NANEUSM20PT.
SQ SEQUENCE 768 AA: 85132 MW: D25DR32 CRC32:
      1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
Query Match 33.9% Score 77: DB 5: Length 768;
Best Local Similarity 27.3% Pred. No. 9,96e-01;
Matches 5: Conservative 13: Mismatches 5: Indels 0:
ID 669 FLVIPGFAIAAVNTRRMNPI 689
      1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
QY 3 YLII SARDVLAIVYVKRMRKVIVAFNFGP 24

```

Search completed: Mon Jun 19 16:11:45 2000  
Job time : 21 secs.

\*\*\*\*\*  
 W E B S I D E  
 (TM)  
 \*\*\*\*\*

Release 3.1A John F. Collins, BioComputing Research Unit.  
 Copyright (C) 1993-1998 University of Edinburgh, U.K.  
 Distribution rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run On: Mon Jun 19 15:14:13 2000; MasPar time 4.00 seconds  
 118,427 Million cell updates/sec  
 Iterular output not generated.

Title: US-09-142-524A-11  
 Description: (1-20) from US09142524A.pep  
 Perfect Score: 143  
 Sequence: IFSKNLNLIKLNPLYAGNK 20

Scoring table: BAW 150  
 Gap 15

Searched: 188463 seqs, 23686106 residues

Post-processing: Minimum Match 0%  
 Listing first 45 summaries

Database: a-geneseq35  
 l:geneseqp

Statistics: Mean 20.352; Variance 65.008; scal: 0.308

Pred. No. is the number of results predicted by chance to have a  
 score greater than or equal to the score of the result being printed,  
 and is derived by analysis of the total score distribution.

# SUMMARIES

| Result No. | Score | Query Match | Length | ID       | Description             | Pred. No. |
|------------|-------|-------------|--------|----------|-------------------------|-----------|
| 1          | 143   | 100.0       | 20     | 1 W42130 | I-cell epitope peptide  | 3.71e-08  |
| 2          | 143   | 100.0       | 20     | 1 W27372 | Multi-epitope peptide   | 3.71e-08  |
| 3          | 143   | 100.0       | 354    | 1 W04344 | Chamaecyparis obtusa p  | 3.71e-08  |
| 4          | 143   | 100.0       | 354    | 1 W42121 | Japanese cypress pollen | 3.71e-08  |
| 5          | 143   | 100.0       | 375    | 1 W04345 | Chamaecyparis obtusa p  | 3.71e-08  |
| 6          | 123   | 86.0        | 367    | 1 R45577 | Jun s I.                | 8.16e-06  |
| 7          | 123   | 86.0        | 370    | 1 R45578 | Jun v I.                | 8.16e-06  |
| 8          | 113   | 79.0        | 20     | 1 R82498 | Cry j I Japanese Cedar  | 1.15e-04  |
| 9          | 113   | 79.0        | 20     | 1 R45549 | Cry j I Pollen allergen | 1.15e-04  |
| 10         | 113   | 79.0        | 30     | 1 W4582  | I-cell epitope peptide  | 1.15e-04  |
| 11         | 113   | 79.0        | 234    | 1 W27371 | Multi-epitope peptide   | 1.15e-04  |
| 12         | 113   | 79.0        | 353    | 1 R75388 | Japanese cedar pollen   | 1.15e-04  |
| 13         | 113   | 79.0        | 353    | 1 R81587 | Cedar pollen allergen   | 1.15e-04  |
| 14         | 113   | 79.0        | 374    | 1 R31937 | Cry j I.                | 1.15e-04  |
| 15         | 113   | 79.0        | 374    | 1 R45541 | Cry j I pollen allergen | 1.15e-04  |
| 16         | 113   | 79.0        | 374    | 1 R82490 | Cry j I Japanese Cedar  | 1.15e-04  |
| 17         | 113   | 79.0        | 374    | 1 R83166 | Japanese cedar pollen   | 1.15e-04  |
| 18         | 74    | 51.7        | 23     | 1 W42131 | I-cell epitope peptide  | 1.15e-04  |
| 19         | 72    | 50.3        | 205    | 1 W27370 | Multi-epitope peptide   | 3.65e+00  |
| 20         | 69    | 48.3        | 23     | 1 W42129 | I-cell epitope peptide  | 7.44e+00  |
| 21         | 67    | 46.9        | 183    | 1 W38604 | S. pneumoniae hemolysin | 1.19e-01  |
| 22         | 64    | 44.8        | 383    | 1 R07347 | Amb a II/Antigen E enc  | 2.39e-01  |
| 23         | 64    | 44.8        | 383    | 1 R07346 | Amb a I/Antigen E enc   | 2.39e-01  |

|    |    |      |     |          |                         |          |
|----|----|------|-----|----------|-------------------------|----------|
| 24 | 64 | 44.8 | 383 | 1 W58730 | Allergen Amb a 10 clon  | 2.39e-01 |
| 25 | 64 | 44.8 | 383 | 1 W41854 | Ragweed pollen Amb a 1  | 2.39e-01 |
| 26 | 64 | 44.8 | 392 | 1 R43587 | Ragweed Pollen Allerge  | 2.39e-01 |
| 27 | 64 | 44.8 | 392 | 1 W02508 | Ragweed Amb a1.4 aller  | 2.39e-01 |
| 28 | 53 | 44.1 | 150 | 1 W08433 | H. pylori GHPO 450 pro  | 3.01e-01 |
| 29 | 62 | 43.4 | 129 | 1 W77573 | Staphylococcus aureus   | 3.78e-01 |
| 30 | 62 | 43.4 | 245 | 1 W64220 | Human secreted protein  | 3.78e-01 |
| 31 | 62 | 43.4 | 387 | 1 R07333 | Amb a I/Antigen E enc   | 3.78e-01 |
| 32 | 62 | 43.4 | 387 | 1 W58727 | Allergen Amb a 1A clon  | 3.78e-01 |
| 33 | 62 | 43.4 | 387 | 1 W41851 | Ragweed pollen Amb a 1  | 3.78e-01 |
| 34 | 62 | 43.4 | 391 | 1 R07327 | Part of Amb a 1/Antige  | 3.78e-01 |
| 35 | 62 | 43.4 | 391 | 1 W58751 | Ragweed allergen UNC C  | 3.78e-01 |
| 36 | 62 | 43.4 | 396 | 1 R43584 | Ragweed Pollen Allerge  | 3.78e-01 |
| 37 | 62 | 43.4 | 396 | 1 W02505 | Ragweed Amb a1.1 aller  | 3.78e-01 |
| 38 | 62 | 43.4 | 398 | 1 R07334 | Amb a I/Antigen E enc   | 3.78e-01 |
| 39 | 62 | 43.4 | 398 | 1 W02506 | Ragweed Amb a1.2 aller  | 3.78e-01 |
| 40 | 62 | 43.4 | 398 | 1 W58729 | Allergen Amb a 1B clon  | 3.78e-01 |
| 41 | 62 | 43.4 | 398 | 1 R43585 | Ragweed Pollen Allerge  | 3.78e-01 |
| 42 | 62 | 43.4 | 440 | 1 W41852 | Ragweed pollen Amb a 1  | 3.78e-01 |
| 43 | 62 | 43.4 | 442 | 1 W41856 | Ragweed pollen UNC C10  | 3.78e-01 |
| 44 | 50 | 42.0 | 275 | 1 W97685 | Staphylococcus aureus   | 5.96e-01 |
| 45 | 50 | 41.3 | 431 | 1 V11853 | Ragweed pollen: Amb a 1 | 7.47e-01 |

## ALIGNMENTS

RESULT 1  
 ID W42130 standard: peptide: 20 AA.  
 AC W42130:  
 DT 16-JUN-1998 (first entry)  
 DE I-cell epitope peptide 10 from Japanese cypress pollen antigen Chao1.  
 KW Japanese cypress pollen; antigen: T-cell epitope; Chao1; Chao2;  
 OS Chamaecyparis obtusa.  
 PN W09747648-A1.  
 PD 18-DEC-1997.  
 PF 12-JUN-1997; J02031.  
 PR 14-JUN-1996; JP-153527.  
 PA (MEIP ) MEIJI MILK PROD CO LTD.  
 PI Dairiki K, Kimo K;  
 DR WPI: 98-052242/G5.  
 PT T-cell epitope peptide portion of Japanese cypress pollen antigens  
 PT Chao1 and Chao2 - used for diagnosis and treatment of spring tree  
 PT pollen disease  
 PS Claim 1: Page 22, 71pp; Japanese.  
 CC The present sequence represents a T-cell epitope peptide from Japanese  
 CC cypress pollen antigen Chao1. The present invention describes peptides  
 CC which correspond to the T-cell epitope sites on Japanese cypress pollen  
 CC antigens Chao1 and Chao2. The peptides can be used as a reagent for the  
 CC diagnosis of allergy to Japanese cypress pollen, and as an antigen in  
 CC the treatment and prevention of spring tree pollen disease in which the  
 CC pollinosis involves reactivity to Japanese cypress pollen.  
 SQ Sequence 20 AA:  
 Query Match: 100.0%; Score 143; DB 1; Length 20;  
 Best Local Similarity 100.0%; Pred. No. 3.71e-08;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 IFSKNLNLIKLNPLYAGNK 20  
 QY 1 IFSKNLNLIKLNPLYAGNK 20

RESULT 2  
 ID W27372 standard: peptide: 31 AA.  
 AC W27372:

DT 24-MAR-1998 (first entry)  
 DE Multi-epitope peptide used as immunotherapeutic agent #4.  
 KW Multi-epitope peptide: immunotherapeutic agent; allergic disease;  
 KW T-cell epitope region; allergen; lymphocyte; immunoglobulin E.  
 OS Synthetic.  
 PN W09712600-A1.  
 PD 12-SEP-1997.





Query Match: 79.0%, Score 113, DB 1, Length 20;  
 Best Local Similarity: 75.0%, Pred. No. 1, 15e-04;  
 Matches: 15; Conservative: 3; Mismatches: 2; Indels: 0; Gaps: 0;

DB 1 IFSGNMNIKLKMPWYIAGYK 20  
 QY 1 IFSKNLNKLKMPWYIAGNK 20

RESULT 10  
 ID W44682 standard; peptide: 30 AA.  
 AC W44682;  
 DT 01-MAY-1998 (first entry)  
 DE T-cell epitope peptide #1 of sugi pollen antigen.  
 KW T-cell epitope: sugi pollen antigen; sugi pollinosis.  
 OS Synthetic.  
 PS Cryptomeria japonica.  
 PR J1000700-A.  
 PD 13-JAN-1998.  
 PE 24-JUN-1996; JP-163287.  
 PA (DAIIC) DAICHI CHEM IND LTD.  
 RA (MEIJI) MEIJI SEIKA KAISHA LTD.  
 RK WPI: 98-133630/13.  
 FT T cell epitope peptide of sugi pollen antigen - useful in the treatment of sugi pollinosis.  
 PS Claim 1; Page 1; 14pp; Japanese.  
 CC T-cell epitope peptides W44682-88 and their derivatives react with sugi pollinosis patient peripheral blood T lymphocytes. A composition prepared by combining at least 2 of the above peptides and/or their derivatives is used for the prevention and treatment of sugi pollinosis.  
 SC Sequence: 30 AA.

Query Match: 79.0%, Score 113, DB 1, Length 30;  
 Best Local Similarity: 75.0%, Pred. No. 1, 15e-04;  
 Matches: 15; Conservative: 3; Mismatches: 2; Indels: 0; Gaps: 0;

DB 1 IFSGNMNIKLKMPWYIAGYK 30  
 QY 1 IFSKNLNKLKMPWYIAGNK 20

RESULT 11  
 ID W27371 standard; peptide: 134 AA.  
 AC W27371;  
 DT 24-MAR-1999 (first entry)  
 DE Multi-epitope peptide used as immunotherapeutic agent #3.  
 KW Multi-epitope peptide: immunotherapeutic agent; allergic diseases; T-cell epitope region; allergen; lymphocyte; immunoglobulin E.  
 OS Synthetic.  
 PR W09732600-A1.  
 PD 12-SEP-1997.  
 PE 10-MAR-1997; J00740.  
 PA (MEIJI) MEIJI MILK PROD CO LTD.  
 RA Dairiki K. Iwara A. Kino K. Kume A. Sone T.  
 RK WPI: 97-470495/43.  
 FT Peptide immunotherapeutic agent to treat allergic diseases - contains multi-epitope peptide containing T cell epitope regions from different allergens.  
 PS Claim 5; Page 32; 58pp; Japanese.  
 CC The present sequence represents a multi-epitope peptide which is used as a new immunotherapeutic agent. It comprises T cell epitope regions from 2 or more different allergens (preferably linked via arginine or lysine dimers), where the T cell epitope regions have a positivity index greater than 100 as measured in a patient group responding to the allergen; have at least 70% reactivity with lymphocytes from patients responding to the allergen; and are not reactive with immunoglobulin E (IgE) antibodies from patients responsive to the allergen. The agent can be used to prevent and treat a wide variety of allergic diseases, e.g. by desensitisation. Side effects, e.g. those mediated by IgE, are reduced.  
 SC Sequence: 134 AA.

Query Match: 79.0%, Score 113, DB 1, Length 134;  
 Best Local Similarity: 75.0%, Pred. No. 1, 15e-04;  
 Matches: 15; Conservative: 3; Mismatches: 2; Indels: 0; Gaps: 0;

DB 78 IFSGNMNIKLKMPWYIAGYK 97  
 QY 1 IFSKNLNKLKMPWYIAGNK 20

RESULT 12  
 ID R75388 standard; protein: 353 AA.  
 AC R75388;  
 DT 12-MAR-1996 (first entry)  
 DE Japanese cedar pollen allergen  
 KW Japanese cedar; pollen allergen; Cryj 1; T-cell epitope; peptides.  
 OS Prevention; treatment; cryptomeria pollinosis.  
 PR Cryptomeria japonica.  
 RA Key Location/Qualifiers  
 FT peptide 61..175  
 FT peptide /note="T-cell epitope peptide"  
 FT peptide 91..105  
 FT peptide /note="T-cell epitope peptide"  
 FT peptide 106..120  
 FT peptide /note="T-cell epitope peptide"  
 FT peptide 146..160  
 FT peptide /note="T-cell epitope peptide"  
 FT peptide 211..225  
 FT peptide /note="T-cell epitope peptide"  
 FT peptide 326..340  
 FT peptide /note="T-cell epitope peptide"  
 FT peptide 335..346  
 FT peptide /note="T-cell epitope peptide"  
 PN J07118295-A.  
 PD 09-MAY-1995.  
 PE 20-OCT-1993; 262626.  
 PR 20-OCT-1993; JP-262626.  
 PA (MEIJI) MEIJI MILK PROD CO LTD.  
 DR WPI: 95-203834/27.  
 FT New cryptomeria pollen allergen T-cell epitope peptide - used for prevention, treatment and investigation of Japanese cedar pollinosis.  
 PS Disclosure: Figs 1-2; App; Japanese.  
 CC R75388 is the Japanese cedar pollen allergen Cryj 1, from which the T-cell epitope peptides R892-R89295 were derived. The peptides can be used for the prevention and treatment of cryptomeria pollinosis, and also for the investigation of pollinosis.  
 SC Sequence: 453 AA.

Query Match: 79.0%, Score 113, DB 1, Length 153;  
 Best Local Similarity: 75.0%, Pred. No. 1, 15e-04;  
 Matches: 15; Conservative: 3; Mismatches: 2; Indels: 0; Gaps: 0;

DB 71 IFSGNMNIKLKMPWYIAGYK 90  
 QY 1 IFSKNLNKLKMPWYIAGNK 20

RESULT 13  
 ID R81587 standard; protein: 353 AA.  
 AC R81587;  
 DT 24-MAY-1996 (first entry)  
 DE Cedar pollen allergen B.  
 KW Cedar; pollen; allergen; immunoglobulin E; IgE; T-cell epitope; antibody; pollinosis; therapy; immunotherapy.  
 OS Cryptomeria japonica.  
 PR EP-700929-A2.  
 PD 13-MAR-1996.  
 PE 08-SEP-1995; 306295.  
 PR 10-SEP-1994; JP-242137.  
 PR 14-JUL-1995; JP-200221.  
 PR 14-JUL-1995; JP-200204.  
 PA (HAYB) HAYASHIBARA SEIBUTSU KAKAKU.  
 PI Hino K, Saito S, Taniguchi Y.  
 DR WPI: 96-140976/15.







CC LENGTH: 383 amino acids  
 CC TYPE: amino acid  
 CC TOPOLOGY: linear  
 CC MOLECULE TYPE: protein  
 CC SEQUENCE 383 AA: 41852 MW: 738971 CN:

Query Match 44.8% Score 64; DB 1; Length 383;  
 Best Local Similarity 35.0%; Pred. No. 1.13e+01;  
 Matches 7; Conservative 6; Mismatches 7; Indels 0; Gaps 0;

DB 110 IFERDMVIRLDRELAINNOK 129

QY 1 IFSKNLNIKLMPPLYIAGNK 20

RESULT 2  
 ID US-08-240-448A-78 STANDARD: PRT: 383 AA.

XX

AC xxxxxx

DT

XX

Sequence 78, Application US/08290448A

Sequence 78, Application US/08290448A

Patent No. 5676954

GENERAL INFORMATION:

APPLICANT: Rogers, Bruce

APPLICANT: Klapper, David G.

APPLICANT: Raina, Thorunn

APPLICANT: Kuo, Mei-chang

TITLE OF INVENTION: Allergic Proteins From Ragweed and Uses

NUMBER OF SEQUENCES: 93

CORRESPONDENCE ADDRESS:

STREET: 60 State Street, suite 510

CITY: Boston

STATE: Massachusetts

COUNTRY: USA

ZIP: 02109-1875

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/78/290 448A

FILING DATE: August 15, 1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/524,951

FILING DATE: May 29, 1992

APPLICATION NUMBER: US 07/325,365

FILING DATE: March 17, 1989

ATTORNEY/AGENT INFORMATION:

NAME: Atty E. Mandragoras

REGISTRATION NUMBER: 36,207

REFERENCE/DOCKET NUMBER: 2M-C18CN

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617)227-7400

TELEFAX: (617)227-5941

INFORMATION FOR SEQ ID NO: 78:

SEQUENCE CHARACTERISTICS:

LENGTH: 383 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE 383 AA: 41852 MW: 738971 CN:

Query Match 44.8% Score 64; DB 1; Length 383;  
 Best Local Similarity 35.0%; Pred. No. 1.13e+01;  
 Matches 7; Conservative 6; Mismatches 7; Indels 0; Gaps 0;

DB 110 IFERDMVIRLDRELAINNOK 129

QY 1 IFSKNLNIKLMPPLYIAGNK 20

RESULT 3

ID US-08-290-448A-78 STANDARD: PRT: 383 AA.

XX

AC xxxxxx

DT

XX

Sequence 78, Application US/08290448A

Sequence 78, Application US/08290448A

Patent No. 5698204

GENERAL INFORMATION:

APPLICANT: Rogers, Bruce

APPLICANT: Klapper, David G.

APPLICANT: Raina, Thorunn

APPLICANT: Kuo, Mei-chang

TITLE OF INVENTION: Allergic Proteins From Ragweed and Uses

NUMBER OF SEQUENCES: 93

CORRESPONDENCE ADDRESS:

STREET: 60 State Street, suite 510

CITY: Boston

STATE: Massachusetts

COUNTRY: USA

ZIP: 02109-1875

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/290 448A

FILING DATE: August 15, 1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/524,951

FILING DATE: May 29, 1992

APPLICATION NUMBER: US 07/325,365

FILING DATE: March 17, 1989

ATTORNEY/AGENT INFORMATION:

NAME: Atty E. Mandragoras

REGISTRATION NUMBER: 36,207

REFERENCE/DOCKET NUMBER: 2M-C18CN

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617)227-7400

TELEFAX: (617)227-5941

INFORMATION FOR SEQ ID NO: 78:

SEQUENCE CHARACTERISTICS:

LENGTH: 383 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE 383 AA: 41852 MW: 738971 CN:

Query Match 44.8% Score 64; DB 1; Length 383;  
 Best Local Similarity 35.0%; Pred. No. 1.13e+01;  
 Matches 7; Conservative 6; Mismatches 7; Indels 0; Gaps 0;

DB 110 IFERDMVIRLDRELAINNOK 129

QY 1 IFSKNLNIKLMPPLYIAGNK 20

RESULT 4

ID US-08-175-069A-72 STANDARD: PRT: 387 AA.

XX

AC xxxxxx

DT

XX



CC SOFTWARE: Patent Release #1.0, Version #1.25  
 CC CURRENT APPLICATION DATA:  
 CC APPLICATION NUMBER: US/08/290.448A  
 CC FILING DATE: August 15, 1994  
 CC PRCR APPLICATION DATA: US 07/529,951  
 CC FILING DATE: May 29, 1990  
 CC APPLICATION NUMBER: US 07/325,365  
 CC FILING DATE: March 17, 1989  
 CC ATTORNEY/AGENT INFORMATION:  
 CC NAME: Amy E. Mandragoras  
 CC REGISTRATION NUMBER: 36,207  
 CC REFERENCE/DOCKET NUMBER: IMI-018CN  
 CC TELECOMMUNICATION INFORMATION:  
 CC TELEPHONE: (617)227-7400  
 CC TELEFAX: (617)227-5941  
 CC INFORMATION FOR SEQ ID NO: 1:  
 CC SEQUENCE CHARACTERISTICS:  
 CC LENGTH: 387 amino acids  
 CC TYPE: amino acid  
 CC TOPOLOGY: linear  
 CC MOLECULE TYPE: protein  
 CC SEQUENCE: 387 AA: 4663 MW: 736367 CN:  
 CC  
 CC Query Match 43.4% Score 62: DB 1: Length 387:  
 CC Best Local Similarity 25.0% Pred. No. 1:82e-01:  
 CC Matches 5: Conservative 10: Mismatches 5: Indels 0: Gaps 0:  
 CC  
 CC 109 IFSKNNVIRLNQELVNSCK 128  
 CC 1 IFSKNNVIRLNQELVNSCK 128  
 CC  
 CC RESULT 7  
 CC US-08-290-448A-59 STANDARD: PRT: 391 AA:  
 CC  
 CC AC xxxxxx  
 CC  
 CC DE Sequence 59, Application US/08290448A  
 CC Sequence 59, Application US/08290448A  
 CC Patent No. 567954  
 CC GENERAL INFORMATION:  
 CC APPLICANT: Rogers, Bruce  
 CC APPLICANT: Klapper, David G.  
 CC APPLICANT: Rainard, Thorunn  
 CC APPLICANT: Kuo, Mei-Chang  
 CC TITLE OF INVENTION: Allogeneic Proteins From Fused and Uses  
 CC NUMBER OF SEQUENCES: 9  
 CC CORRESPONDENCE ADDRESS:  
 CC ADDRESSEE: LAHIVE & COCKFIELD  
 CC STREET: 60 State Street, suite 510  
 CC CITY: Boston  
 CC STATE: Massachusetts  
 CC COUNTRY: USA  
 CC ZIP: 02109-1875  
 CC COMPUTER READABLE FORM:  
 CC MEDIUM TYPE: floppy disk  
 CC COMPUTER: IBM PC compatible  
 CC OPERATING SYSTEM: PC-DOS/MS-DOS  
 CC SOFTWARE: Patent Release #1.0, Version #1.25  
 CC CURRENT APPLICATION DATA:  
 CC APPLICATION NUMBER: US/08/290.448A  
 CC FILING DATE: August 15, 1994  
 CC PRCR APPLICATION DATA:  
 CC APPLICATION NUMBER: US 07/529,951  
 CC FILING DATE: May 29, 1990  
 CC APPLICATION NUMBER: US 07/325,365  
 CC FILING DATE: March 17, 1989  
 CC ATTORNEY/AGENT INFORMATION:  
 CC NAME: Amy E. Mandragoras

CC REGISTRATION NUMBER: 36,207  
 CC REFERENCE/DOCKET NUMBER: IMI-018CN  
 CC TELECOMMUNICATION INFORMATION:  
 CC TELEPHONE: (617)227-7400  
 CC TELEFAX: (617)227-5941  
 CC INFORMATION FOR SEQ ID NO: 59:  
 CC SEQUENCE CHARACTERISTICS:  
 CC LENGTH: 391 amino acids  
 CC TYPE: amino acid  
 CC TOPOLOGY: linear  
 CC MOLECULE TYPE: protein  
 CC SEQUENCE: 391 AA: 4264 MW: 754045 CN:  
 CC  
 CC Query Match 43.4% Score 62: DB 1: Length 391:  
 CC Best Local Similarity 40.0% Pred. No. 1:80e-01:  
 CC Matches 8: Conservative 5: Mismatches 7: Indels 0: Gaps 0:  
 CC  
 CC 113 IFSKNNVIRLNQELVNSCK 132  
 CC 1 IFSKNNVIRLNQELVNSCK 132  
 CC  
 CC RESULT 8  
 CC US-08-290-448A-59 STANDARD: PRT: 391 AA:  
 CC  
 CC AC xxxxxx  
 CC  
 CC DE Sequence 59, Application US/08290448A  
 CC Sequence 59, Application US/08290448A  
 CC Patent No. 5698204  
 CC GENERAL INFORMATION:  
 CC APPLICANT: Rogers, Bruce  
 CC APPLICANT: Klapper, David G.  
 CC APPLICANT: Rainard, Thorunn  
 CC APPLICANT: Kuo, Mei-Chang  
 CC TITLE OF INVENTION: Allogeneic Proteins From Fused and Uses  
 CC NUMBER OF SEQUENCES: 93  
 CC CORRESPONDENCE ADDRESS:  
 CC ADDRESSEE: LAHIVE & COCKFIELD  
 CC STREET: 60 State Street, suite 510  
 CC CITY: Boston  
 CC STATE: Massachusetts  
 CC COUNTRY: USA  
 CC ZIP: 02109-1875  
 CC COMPUTER READABLE FORM:  
 CC MEDIUM TYPE: floppy disk  
 CC COMPUTER: IBM PC compatible  
 CC OPERATING SYSTEM: PC-DOS/MS-DOS  
 CC SOFTWARE: Patent Release #1.0, Version #1.25  
 CC CURRENT APPLICATION DATA:  
 CC APPLICATION NUMBER: US/08/290.448A  
 CC FILING DATE: August 15, 1994  
 CC PRCR APPLICATION DATA:  
 CC APPLICATION NUMBER: US 07/529,951  
 CC FILING DATE: May 29, 1990  
 CC APPLICATION NUMBER: US 07/325,365  
 CC FILING DATE: March 17, 1989  
 CC ATTORNEY/AGENT INFORMATION:  
 CC NAME: Amy E. Mandragoras  
 CC REGISTRATION NUMBER: 36,207  
 CC REFERENCE/DOCKET NUMBER: IMI-018CN  
 CC TELECOMMUNICATION INFORMATION:  
 CC TELEPHONE: (617)227-7400  
 CC TELEFAX: (617)227-5941  
 CC INFORMATION FOR SEQ ID NO: 59:  
 CC SEQUENCE CHARACTERISTICS:  
 CC LENGTH: 391 amino acids  
 CC TYPE: amino acid  
 CC TOPOLOGY: linear  
 CC MOLECULE TYPE: protein







Matches 7: Conservative 7: Mismatches 6: Indels 0: Gaps 0:

00 119 11KNDVV:NLNDELWVNSDK 138

01 119 11KNDVV:NLNDELWVNSDK 138

02 119 11KNDVV:NLNDELWVNSDK 138

Search completed: Tue Jun 20 14:36:45 2000  
Job time : 5 sec.







CC FILING DATE: April 8, 1994  
CC APPLICATION NUMBER: 07/938,990  
CC FILING DATE: September 1, 1992  
CC APPLICATION NUMBER: PCT/US93/00139  
CC FILING DATE: January 15, 1993  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Darlene A. Vanstone  
CC REGISTRATION NUMBER: 35,729  
CC REFERENCE/DOCKET NUMBER: 025.6 US (IMI-028CP2)  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (517) 466-6000  
CC TELEFAX: (517) 466-6040  
CC INFORMATION FOR SEQ ID NO: 95:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 367 amino acids  
CC TYPE: amino acid  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: protein  
CC SEQUENCE 367 AA: 39846 MW: 665078 CN:  
SQ  
Query Match 86.0% Score 123, DB 10, Length 367,  
Best Local Similarity 75.0%, Pred.No. 1,25e-05,  
Matches 15: Conservative 4: Mismatches 1: Indels 0: Gaps 0:  
DB 92 IFSONMNIKLMPLYVAGHK 111  
QY 1 IFSKNLNIKLMPLYIAGNK 20  
RESULT 5  
ID US-08-226-248A-95 STANDARD: PRT: 367 AA.  
AC xxxxxx  
AT  
CT  
DT  
ET  
FF  
Sequence 95, Application US/08226248A  
Sequence 95, Application US/08226248A  
GENERAL INFORMATION:  
APPLICANT: Griffith, Irwin J.  
APPLICANT: Pollock, Joanne  
APPLICANT: Bond, Julian F.  
APPLICANT: Garman, Richard D.  
APPLICANT: Kuo, Mei-Chang  
APPLICANT: Yeung, Siu-Mei H.  
APPLICANT: Brauer, Andrew  
APPLICANT: Exley, Mark A.  
APPLICANT: Powers, Steven P.  
TITLE OF INVENTION: Allergenic Proteins And Peptides From  
TITLE OF INVENTION: Japanese Cedar Pollen  
NUMBER OF SEQUENCES: 201  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Immunologic Pharmaceutical Corporation, Inc.  
STREET: 610 Lincoln St  
CITY: Waltham  
STATE: MA  
COUNTRY: USA  
ZIP: 02154  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/226,248A  
FILING DATE: April 8, 1994  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/938,990  
FILING DATE: September 1, 1992  
APPLICATION NUMBER: PCT/US93/00139  
FILING DATE: January 15, 1993

CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Darlene A. Vanstone  
CC REGISTRATION NUMBER: 35,729  
CC REFERENCE/DOCKET NUMBER: 025.5 US (IMI-028CP)  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (517) 466-6000  
CC TELEFAX: (517) 466-6040  
CC INFORMATION FOR SEQ ID NO: 95:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 367 amino acids  
CC TYPE: amino acid  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: protein  
CC SEQUENCE 367 AA: 39846 MW: 665078 CN:  
SQ  
Query Match 86.0% Score 123, DB 7, Length 367,  
Best Local Similarity 75.0%, Pred.No. 1,25e-05,  
Matches 15: Conservative 4: Mismatches 1: Indels 0: Gaps 0:  
DB 92 IFSONMNIKLMPLYVAGHK 111  
QY 1 IFSKNLNIKLMPLYIAGNK 20  
RESULT 6  
ID US-08-350-225-95 STANDARD: PRT: 367 AA.  
AC xxxxxx  
AT  
CT  
DT  
ET  
FF  
Sequence 95, Application US/08350225  
Sequence 95, Application US/08350225  
GENERAL INFORMATION:  
APPLICANT: Griffith, Irwin J.  
APPLICANT: Pollock, Joanne  
APPLICANT: Bond, Julian F.  
APPLICANT: Garman, Richard D.  
APPLICANT: Kuo, Mei-Chang  
APPLICANT: Yeung, Siu-Mei H.  
APPLICANT: Brauer, Andrew  
APPLICANT: Exley, Mark A.  
APPLICANT: Powers, Steven P.  
TITLE OF INVENTION: Allergenic Proteins And Peptides From  
TITLE OF INVENTION: Japanese Cedar Pollen  
NUMBER OF SEQUENCES: 261  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Immunologic Pharmaceutical Corporation, Inc.  
STREET: 610 Lincoln St  
CITY: Waltham  
STATE: MA  
COUNTRY: USA  
ZIP: 02154  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/350,225  
FILING DATE: December 6, 1994  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/226,248  
FILING DATE: April 8, 1994  
APPLICATION NUMBER: 07/938,990  
FILING DATE: September 1, 1992  
APPLICATION NUMBER: PCT/US93/00139  
FILING DATE: January 15, 1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Darlene A. Vanstone  
REGISTRATION NUMBER: 35,729



Matches 15: Conservative 4: Mismatches 1: Indels 0: Gaps 0:

DB 92 IFSQNNIKLNKMPYVAGHK 111  
IIIIIIIIIIIIIIIIIIII  
QY 1 IFSKNNIKLNKMPYIAGNK 20

RESULT 9  
ID US-08-467-021-47 STANDARD: PRI: 370 AA.  
XX  
AC xxxxxx  
XX  
DT  
XX

Sequence 97, Application US/08457023

Sequence 97, Application US/08457023

GENERAL INFORMATION:  
APPLICANT: Griffith, Irwin J.  
APPLICANT: Pollock, Joanne  
APPLICANT: Bond, Julian F.  
APPLICANT: Garman, Richard D.  
APPLICANT: Kuo, Mei-Chang  
APPLICANT: Yeung, Siu-mei H.  
APPLICANT: Brauer, Andrew  
APPLICANT: Exley, Mark A.  
APPLICANT: Powers, Steven P.  
TITLE OF INVENTION: Allergenic Proteins And Peptides From  
TITLE OF INVENTION: Japanese Cedar Pollen  
NUMBER OF SEQUENCES: 261  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: ImmuLogic Pharmaceutical Corporation, Inc.  
STREET: 610 Lincoln St  
CITY: Waltham  
STATE: MA  
COUNTRY: USA  
ZIP: 02154

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION NUMBER: US/08/467.023  
FILING DATE: June 6, 1995  
PRIORITY APPLICATION DATA:  
CLASSIFICATION: 424  
APPLICATION NUMBER: 08/350.225  
FILING DATE: December 5, 1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Jane E. Remillard  
REGISTRATION NUMBER: 38,872  
REFERENCE/DOCKET NUMBER: 025.6 USD2 (IMI-028CPD2)  
TELEPHONE: (617) 227-7400  
TELEFAX: (617) 227-5941  
INFORMATION FOR SEQ ID NO: 97:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 370 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE 370 AA: 40191 MW: 725506 CN:

Query Match 86.0% Score 123; DB 10; Length 370;  
Best Local Similarity 75.0%; Pred. No. 1.26e-05;  
Matches 15; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

DB 92 IFSQNNIKLNKMPYVAGHK 111  
IIIIIIIIIIIIIIIIIIII  
QY 1 IFSKNNIKLNKMPYIAGNK 20

RESULT 10  
ID US-08-467-697-97 STANDARD: PRI: 370 AA.  
XX  
AC xxxxxx  
XX  
DT  
XX

Sequence 97, Application US/08457697

Sequence 97, Application US/08457697

GENERAL INFORMATION:  
APPLICANT: Griffith, Irwin J.  
APPLICANT: Pollock, Joanne  
APPLICANT: Bond, Julian F.  
APPLICANT: Garman, Richard D.  
APPLICANT: Kuo, Mei-Chang  
APPLICANT: Yeung, Siu-mei H.  
APPLICANT: Brauer, Andrew  
APPLICANT: Exley, Mark A.  
APPLICANT: Powers, Steven P.  
TITLE OF INVENTION: Allergenic Proteins And Peptides From  
TITLE OF INVENTION: Japanese Cedar Pollen  
NUMBER OF SEQUENCES: 261  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: ImmuLogic Pharmaceutical Corporation, Inc.  
STREET: 610 Lincoln St  
CITY: Waltham  
STATE: MA  
COUNTRY: USA  
ZIP: 02154

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION NUMBER: US/08/467.697  
FILING DATE: June 6, 1995  
PRIORITY APPLICATION DATA:  
CLASSIFICATION: 424  
APPLICATION NUMBER: 350.225  
FILING DATE: December 5, 1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Jane E. Remillard  
REGISTRATION NUMBER: 38,872  
REFERENCE/DOCKET NUMBER: 025.6 USD4 (IMI-028704)  
TELEPHONE: (617) 227-7400  
TELEFAX: (617) 227-5941  
INFORMATION FOR SEQ ID NO: 97:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 370 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE 370 AA: 40191 MW: 725506 CN:

Query Match 86.0% Score 123; DB 10; Length 370;  
Best Local Similarity 75.0%; Pred. No. 1.26e-05;  
Matches 15; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

DB 92 IFSQNNIKLNKMPYVAGHK 111  
IIIIIIIIIIIIIIIIIIII  
QY 1 IFSKNNIKLNKMPYIAGNK 20

RESULT 11  
ID US-08-226-249A-97 STANDARD: PRI: 370 AA.  
XX  
AC xxxxxx  
XX  
DT  
XX

```

DE Sequence 97, Application US/38226248A
XX
XX Sequence 97, Application US/08226248A
XX GENERAL INFORMATION:
XX APPLICANT: Griffith, Irwin J.;
XX APPLICANT: Pollock, Joanne;
XX APPLICANT: Bond, Julian;
XX APPLICANT: Garman, Richard D.;
XX APPLICANT: Kuo, Mei-Chang;
XX APPLICANT: Yeung, Siu-mei H.;
XX APPLICANT: Brauer, Andrew;
XX APPLICANT: Exley, Mark A.;
XX APPLICANT: Powers, Steven P.;
XX TITLE OF INVENTION: Allergic Proteins And Peptides From
XX TITLE OF INVENTION: Japanese Cedar Pollen
XX NUMBER OF SEQUENCES: 20
XX CORRESPONDENCE ADDRESS:
XX ADDRESSEE: Immunologic Pharmaceutical Corporation, Inc.
XX STREET: 610 Lincoln St.
XX CITY: Waltham
XX STATE: MA
XX COUNTRY: USA
XX ZIP: 02154
XX COMPUTER READABLE FORM:
XX MEDIUM TYPE: Floppy disk
XX COMPUTER: IBM PC compatible
XX OPERATING SYSTEM: PC-DOS/MS-DOS
XX SOFTWARE: Patent In Release #1.0, Version #1.25
XX CURRENT APPLICATION DATA:
XX APPLICATION NUMBER: US/08/226,248A
XX FILING DATE: April 8, 1994
XX CLASSIFICATION: 424
XX PRIOR APPLICATION DATA:
XX APPLICATION NUMBER: 07/938,930
XX FILING DATE: September 1, 1992
XX APPLICATION NUMBER: PCT/US93/00139
XX FILING DATE: January 15, 1993
XX ATTORNEY/AGENT INFORMATION:
XX NAME: Darlene A. Varstone
XX REGISTRATION NUMBER: 35,729
XX REFERENCE/DOCKET NUMBER: 025.5 US (IM1-028CP);
XX TELECOMMUNICATION INFORMATION:
XX TELEPHONE: (617) 466-6000
XX TELEFAX: (617) 466-6040
XX INFORMATION FOR SEQ ID NO: 97:
XX SEQUENCE CHARACTERISTICS:
XX LENGTH: 370 amino acids
XX TYPE: amino acid
XX TOPOLOGY: linear
XX MOLECULE TYPE: protein
XX SEQUENCE: 370 AA; 4319 MW; 725506 CN;

Query Match 86.0% Score 123; DB 7; Length 370;
Best Local Similarity 75.0% Pred. No. 1.26e-05;
Matches 15; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 92 IFSONNKKLNKMPLYVAGHK 111
QY 1 IFSKNLNKLNKMPLYIAGNK 20

RESULT 12
ID US-09-468-940-97 STANDARD: PRT: 370 AA.
XX
XX AC xxxxxx
XX
XX DT
XX
XX DE
XX Sequence 97, Application US/08468940
XX
XX Sequence 97, Application US/08468940
XX GENERAL INFORMATION:
XX APPLICANT: Griffith, Irwin J.;
XX APPLICANT: Pollock, Joanne;
XX APPLICANT: Bond, Julian;
XX APPLICANT: Garman, Richard D.;
XX APPLICANT: Kuo, Mei-Chang;
XX APPLICANT: Yeung, Siu-mei H.;
XX APPLICANT: Brauer, Andrew;
XX APPLICANT: Exley, Mark A.;
XX APPLICANT: Powers, Steven P.;
XX TITLE OF INVENTION: Allergic Proteins And Peptides From
XX TITLE OF INVENTION: Japanese Cedar Pollen
XX NUMBER OF SEQUENCES: 20
XX CORRESPONDENCE ADDRESS:
XX ADDRESSEE: Immunologic Pharmaceutical Corporation, Inc.
XX STREET: 610 Lincoln St.
XX CITY: Waltham
XX STATE: MA
XX COUNTRY: USA
XX ZIP: 02154
XX COMPUTER READABLE FORM:
XX MEDIUM TYPE: Floppy disk
XX COMPUTER: IBM PC compatible
XX OPERATING SYSTEM: PC-DOS/MS-DOS
XX SOFTWARE: Patent In Release #1.0, Version #1.25
XX CURRENT APPLICATION DATA:
XX APPLICATION NUMBER: US/08/226,248A
XX FILING DATE: April 8, 1994
XX CLASSIFICATION: 424
XX PRIOR APPLICATION DATA:
XX APPLICATION NUMBER: 08/350,225
XX FILING DATE: December 6, 1994
XX APPLICATION NUMBER: 08/226,248
XX FILING DATE: April 8, 1994
XX APPLICATION NUMBER: 07/938,990
XX FILING DATE: September 1, 1992
XX APPLICATION NUMBER: PCT/US93/00139
XX FILING DATE: January 15, 1994
XX ATTORNEY/AGENT INFORMATION:
XX NAME: Darlene A. Varstone
XX REGISTRATION NUMBER: 35,729
XX REFERENCE/DOCKET NUMBER: 025.4 US (IM1-028CP2);
XX TELECOMMUNICATION INFORMATION:
XX TELEPHONE: (617) 466-6000
XX TELEFAX: (617) 466-6040
XX INFORMATION FOR SEQ ID NO: 97:
XX SEQUENCE CHARACTERISTICS:
XX LENGTH: 370 amino acids
XX TYPE: amino acid
XX TOPOLOGY: linear
XX MOLECULE TYPE: protein
XX SEQUENCE: 370 AA; 4319 MW; 725506 CN;

Query Match 86.0% Score 123; DB 10; Length 370;
Best Local Similarity 75.0% Pred. No. 1.26e-05;
Matches 15; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 92 IFSONNKKLNKMPLYVAGHK 111
QY 1 IFSKNLNKLNKMPLYIAGNK 20

RESULT 13
ID US-09-467-006-97 STANDARD: PRT: 370 AA.
XX
XX AC xxxxxx
XX
XX DT
XX
XX DE
XX Sequence 97, Application US/08467006
XX
XX Sequence 97, Application US/08467006
XX GENERAL INFORMATION:
XX APPLICANT: Griffith, Irwin J.;
XX APPLICANT: Pollock, Joanne;

```

CC APPLICANT: Bond, Julian F.;  
 CC APPLICANT: Garman, Richard D;  
 CC APPLICANT: Kuo, Mei-Chang; H.;  
 CC APPLICANT: Yeung, Siu-mei H.;  
 CC APPLICANT: Brauer, Andrew;  
 CC APPLICANT: Exley, Mark A.;  
 CC APPLICANT: Powers, Steven P.  
 CC TITLE OF INVENTION: Allergenic Proteins And Peptides From  
 CC TITLE OF INVENTION: Japanese Cedar Pollen  
 CC NUMBER OF SEQUENCES: 261  
 CC CORRESPONDENCE ADDRESS:  
 CC ADDRESSEE: Immunologic Pharmaceutical Corporation, Inc.  
 CC STREET: 610 Lincoln St  
 CC CITY: Waltham  
 CC STATE: MA  
 CC COUNTRY: USA  
 CC ZIP: 02154  
 CC COMPUTER READABLE FORM:  
 CC MEDIUM TYPE: Floppy disk  
 CC COMPUTER: IBM PC compatible  
 CC OPERATING SYSTEM: PC-DOS/MS-DOS  
 CC SOFTWARE: PatentIn Release #1.0, Version #1.25  
 CC CURRENT APPLICATION DATA:  
 CC APPLICATION NUMBER: US/08/467,006  
 CC FILING DATE: June 6, 1995  
 CC CLASSIFICATION: 424  
 CC PRIOR APPLICATION DATA:  
 CC APPLICATION NUMBER: 08/350,225  
 CC FILING DATE: December 6, 1994  
 CC ATTORNEY/AGENT INFORMATION:  
 CC NAME: Jane E. Remillard  
 CC REGISTRATION NUMBER: 38,872  
 CC REFERENCE/DOCKET NUMBER: 025.5 US/09 (MI-028CPD6)  
 CC TELECOMMUNICATION INFORMATION:  
 CC TELEPHONE: (617) 227-7400  
 CC TELEFAX: (617) 227-5941  
 CC INFORMATION FOR SEQ ID NO: 97:  
 CC SEQUENCE CHARACTERISTICS:  
 CC LENGTH: 370 amino acids  
 CC TYPE: amino acid  
 CC TOPOLOGY: linear  
 CC MOLECULE TYPE: protein  
 CC SEQUENCE 370 AA: 40191 MW: 725506 CN:  
 CC  
 CC Query Match 86.0% Score 123. PB 8; Length 370;  
 CC Best Local Similarity 75.0% Pred No: 1,24e-05;  
 CC Matches 15: Conservation 4; Mismatches 1; Indels 0; Gaps 0;  
 CC  
 CC 92 IFSNNIKIKKMPLYVAGHK 111  
 CC III III IIIIIII I  
 CC 1 IFSKNLNKLNMPLYAGNK 20  
 CC  
 CC RESULT 14  
 CC ID US-08-350-225-97 STANDARD: PRT: 370 AA.  
 CC XX xxxxxx  
 CC AC  
 CC XX  
 CC DT  
 CC DE  
 CC XX  
 CC Sequence 97, Application US/08350225  
 CC  
 CC Sequence 97, Application US/08350225  
 CC GENERAL INFORMATION:  
 CC APPLICANT: Griffith, Irwin J.;  
 CC APPLICANT: Pollock, Joanne;  
 CC APPLICANT: Bond, Julian F.;  
 CC APPLICANT: Garman, Richard D;  
 CC APPLICANT: Kuo, Mei-Chang;  
 CC APPLICANT: Yeung, Siu-mei H.;  
 CC APPLICANT: Brauer, Andrew;  
 CC APPLICANT: Exley, Mark A.;  
 CC APPLICANT: Powers, Steven P.

CC TITLE OF INVENTION: Allergenic Proteins And Peptides From  
 CC TITLE OF INVENTION: Japanese Cedar Pollen  
 CC NUMBER OF SEQUENCES: 26;  
 CC CORRESPONDENCE ADDRESS:  
 CC ADDRESSEE: Immunologic Pharmaceutical Corporation, Inc  
 CC STREET: 610 Lincoln St  
 CC CITY: Waltham  
 CC STATE: MA  
 CC COUNTRY: USA  
 CC ZIP: 02154  
 CC COMPUTER READABLE FORM:  
 CC MEDIUM TYPE: Floppy disk  
 CC COMPUTER: IBM PC compatible  
 CC OPERATING SYSTEM: PC-DOS/MS-DOS  
 CC SOFTWARE: PatentIn Release #1.0, Version #1.25  
 CC CURRENT APPLICATION DATA:  
 CC APPLICATION NUMBER: US/08/350,225  
 CC FILING DATE: December 6, 1994  
 CC CLASSIFICATION: 424  
 CC PRIOR APPLICATION DATA:  
 CC APPLICATION NUMBER: 08/226,248  
 CC FILING DATE: April 8, 1994  
 CC APPLICATION NUMBER: 07/938,940  
 CC FILING DATE: September 1, 1992  
 CC APPLICATION NUMBER: PCT/US93/00149  
 CC FILING DATE: January 15, 1993  
 CC ATTORNEY/AGENT INFORMATION:  
 CC NAME: Darlene A. Vanstone  
 CC REGISTRATION NUMBER: 35,729  
 CC REFERENCE/DOCKET NUMBER: 025.6 US (MI-028CP2)  
 CC TELECOMMUNICATION INFORMATION:  
 CC TELEPHONE: (617) 466-6000  
 CC TELEFAX: (617) 466-6040  
 CC INFORMATION FOR SEQ ID NO: 97:  
 CC SEQUENCE CHARACTERISTICS:  
 CC LENGTH: 370 amino acids  
 CC TYPE: amino acid  
 CC TOPOLOGY: linear  
 CC MOLECULE TYPE: protein  
 CC SEQUENCE 370 AA: 40191 MW: 745506 CN:  
 CC  
 CC Query Match 86.0% Score 123. PB 8; Length 370;  
 CC Best Local Similarity 75.0% Pred No: 1,24e-05;  
 CC Matches 15: Conservation 4; Mismatches 1; Indels 0; Gaps 0;  
 CC  
 CC 92 IFSNNIKIKKMPLYVAGHK 111  
 CC III III IIIIIII I  
 CC 1 IFSKNLNKLNMPLYAGNK 20  
 CC  
 CC RESULT 15  
 CC ID US-08-467-006-33 STANDARD: PRT: 20 AA.  
 CC XX xxxxxx  
 CC AC  
 CC XX  
 CC DT  
 CC DE  
 CC XX  
 CC Sequence 33, Application US/08467006  
 CC  
 CC Sequence 33, Application US/08467006  
 CC GENERAL INFORMATION:  
 CC APPLICANT: Griffith, Irwin J.;  
 CC APPLICANT: Pollock, Joanne;  
 CC APPLICANT: Bond, Julian F.;  
 CC APPLICANT: Garman, Richard D;  
 CC APPLICANT: Kuo, Mei-Chang;  
 CC APPLICANT: Yeung, Siu-mei H.;  
 CC APPLICANT: Brauer, Andrew;  
 CC APPLICANT: Exley, Mark A.;  
 CC APPLICANT: Powers, Steven P.  
 CC TITLE OF INVENTION: Allergenic Proteins And Peptides From  
 CC TITLE OF INVENTION: Japanese Cedar Pollen  
 CC NUMBER OF SEQUENCES: 261

CC CORRESPONDENCE ADDRESS:  
 CC ADDRESSEE: Immunologic Pharmaceuticals Corporation, Inc.  
 CC STREET: 610 Lincoln St  
 CC CITY: Waltham  
 CC STATE: MA  
 CC COUNTRY: USA  
 CC ZIP: 02154  
 CC COMPUTER READABLE FORM:  
 CC MEDIUM TYPE: Floppy disk  
 CC COMPUTER: IBM PC compatible  
 CC OPERATING SYSTEM: PC-DOS/MS-DOS  
 CC SOFTWARE: Patent Release #1.0, Version #1.25  
 CC CURRENT APPLICATION DATA:  
 CC APPLICATION NUMBER: US/08/467,006  
 CC FILING DATE: June 6, 1995  
 CC CLASSIFICATION: 424  
 CC PRIOR APPLICATION DATA:  
 CC APPLICATION NUMBER: 08/350,225  
 CC FILING DATE: December 6, 1994  
 CC ATTORNEY/AGENT INFORMATION:  
 CC NAME: Jane E. Remillard  
 CC REGISTRATION NUMBER: 38,872  
 CC REFERENCE/DOCKET NUMBER: 025.6 USD6 (IM-028CP06)  
 CC TELECOMMUNICATION INFORMATION:  
 CC TELEPHONE: (617) 227-7400  
 CC TELEFAX: (617) 227-5941  
 CC INFORMATION FOR SEQ ID NO: 33:  
 CC SEQUENCE CHARACTERISTICS:  
 CC LENGTH: 20 amino acids  
 CC TYPE: amino acid  
 CC TOPOLOGY: linear  
 CC MOLECULE TYPE: peptide  
 CC FRAGMENT TYPE: internal  
 CC SEQUENCE 20 AA: 2320 MW: 2402 CN:

Query Match 79.0% Score 113 DB 10 Length 20  
 Best Local Similarity 75.0% Pref No. 2,25e-04  
 Matches 15: Conservative 3: Mismatches 2: Indels 0: Gaps 0:

10 1 TFSNNMKIKLWMPYIAGNK 20  
 11 TFSNNMKIKLWMPYIAGNK 20  
 12 TFSNNMKIKLWMPYIAGNK 20

Search completed: Mon Jun 19 16:27:27 2000  
 Ser time : 19 sec.







```

||||| |||| | | | |
QY 1 IFSKNLNKLNPLYIAGNK 2:

RESULT      8
ENTRY       E64320      *type complete
TITLE       Hypothetical protein MJ0164 - Methanococcus jannaschii
ORGANISM    #formal_name Methanococcus jannaschii
DATE        13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change
              10-Oct-1997
ACCESSIONS  E64320
REFERENCE   A64300
#authors    Bolt, C.J.; White, O.; Olsen, G.J.; Zhou, J.; Fleischmann, R.D.; Sutton, J.B.; Blake, J.A.; FitzGerald, L.M.; Clayton, R.A.; Gocayne, D.; Kerlavage, A.R.; Dougherty, B.A.; Tomb, J.F.; Adams, M.D.; Reich, C.I.; Overbeek, K.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.; Scott, J.L.; Hughes, N.S.M.; Weidman, J.F.; Peterson, J.L.; Nguyen, J.; Utterback, T.R.; Kelley, J.M.; Peterson, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hirst, M.A.; Kane, B.P.; Borodovsky, M.; Klein, R.P.; Fraser, C.M.; Smith, H.G.O.; Woese, C.R.; Venturi, J.C. Science (1996) 273:1158-1173
#journal    Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii.
#accession E64320
#status     preliminary; nucleic acid sequence not shown.
#molecule_type DNA
#residues   1-335 #label BCJL
#cross-references GB:067473; GR:177117; MID:g1590914; PID:g1498914; TIGR:MJ0164

GENETICS
#map_position REV158614-167427
SUMMARY      #length 395 #molecular_weight 46273 #checksum 254
              44.8%; Score 64; DB 2; Length 395;
              Best Local Similarity 64.1%; Pred No. 5,09e+00;
              Matches 9; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Db 129 IEVKYNLKIDPL 142
| 1 |||||::|||
QY 1 IFSKNLNKLNMPYL 14
| 1 |||||::|||

RESULTS      9
ENTRY       E64638      *type complete
TITLE       conserved hypothetical serine protease HP0314 - Helicobacter pylori (strain 26695)
ORGANISM    #formal_name Helicobacter pylori
DATE        09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change
              26-Aug-1999
ACCESSIONS  E64638
REFERENCE   A64520
#authors    Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.; Ketchum, K.A.; Klenk, H.P.; Gill, S.; Dougherty, B.A.; Nelson, K.; Quackenbush, J.; Zhou, L.; Kirkness, E.F.; Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKenney, K.; Fitzgerald, L.M.; Lee, N.; Adams, M.D.; Hickey, E.K.; Berg, D.E.; Gocayne, J.D.; Utterback, J.R.; Peterson, J.L.; Kelley, J.M.; Cotton, M.D.; Weisman, J.M.; Fujii, C.; Bowman, C.; Watthey, L.; Wallin, E.; Hayes, W.S.; Borod'sky, M.; Karpe, P.D.; Smith, H.O.; Fraser, C.M.; Venter, C. Nature (1997) 38 :539-547
#journal    The complete genome sequence of the gastric pathogen Helicobacter pylori.
#cross-references MIM:97394467
#accession E64638
#status     preliminary; nucleic acid sequence not shown.

```



```

Matches      8:  Conservative      4:  Mismatches      2:  Indels      0:  Gaps      0:

DB      5  IFKKILSLINMAL 18
      1  IFSKLNLIKLNMPLE 14
      1  IFSKLNLIKLNMPLE 14

RESULT  14
ENTRY   *type complete
TITLE   *product protocatechuate 3,4-dioxygenase (EC 1.13.11.3) alpha chain -
        *pseudomonas putida
ORGANISM *formal_name Pseudomonas putida
        *note strain ATCC 23975 was formerly classified as Pseudomonas
        aeruginosa
DATE    31-Jan-1990 *sequence_revision 14-Feb-1997 *text_change
        11-Jun-1999
ACCESSIONS B36930: AC0505
REFERENCE  A36930
        *authors Frazee, R.W.; Livingston, D.M.; Laporte, D.C.; Lipscomb, J.D.
        *journal J. Bacteriol. (1993) 175:6194-6202
        *title Cloning, sequencing, and expression of the Pseudomonas putida
        protocatechuate 3,4-dioxygenase genes.
        *cross-references MJD:94012480
        *accession B36930
        *molecule_type DNA
        *residues 1-201 *label FRA
        *cross-references GB:L14836; NID:q294343; PIDN:AA841025.1; PID:q294345
        *experimental_source ATCC 23975

REFERENCE  A00505
        *authors Kohlmeier, N.A.; Howard, J.B.
        *journal C. Biol. Chem. (1979) 254:7309-7315
        *title The primary structure of the alpha subunit of protocatechuate
        3,4-dioxygenase. II. Isolation and sequence of overlap
        peptides and complete sequence.
        *cross-references MJD:79216432
        *accession A00505
        *molecule_type protein
        *residues 2-59, D', 61-76, D', 78-201 *label KOH
        *experimental_source ATCC 23975
        *note this is the second of two papers giving the experimental
        details

REFERENCE  A52897
        *authors Ohlendorf, D.H.; Orville, A.M.; Lipscomb, J.D.
        *submission submitted to the Brookhaven Protein Data Bank, June 1994
        *cross-references PDB:2PCD
        *contents annotation: X-ray crystallography. 2.15 angstroms, residues
        2-201
        *label A58475
        *authors Ohlendorf, D.H.; Orville, A.M.; Lipscomb, J.D.
        *journal J. Mol. Biol. (1994) 244:586-608
        *title Structure of protocatechuate 3,4-dioxygenase from Pseudomonas
        aeruginosa at 2.15 angstroms resolution.
        *cross-references MJD:95082024
        *contents annotation: X-ray crystallography. 2.15 angstroms
        A50309
        *authors Ohlendorf, D.H.; Weber, P.C.
        *submission submitted to the Brookhaven Protein Data Bank, September 1990
        *cross-references PDB:1PCD
        *contents annotation: X-ray crystallography. 2.8 angstroms, residues
        2-24, G', 25-59, D', 61-76, D', 78-135, G', 136-173, 175-201
        *label A58474
        *authors Ohlendorf, D.H.; Lipscomb, J.D.; Weber, P.C.
        *journal Nature (1988) 336:403-405
        *title Structure and assembly of protocatechuate 3,4-dioxygenase.
        *cross-references MJD:89057125
        *contents annotation: X-ray crystallography. 2.8 angstroms

GENETICS
        *gene pcag
        *description dodecamer of heterodimers of alpha and beta chains
        *function catalyzes the oxidative cleavage of 3,4-dihydroxybenzoate to
        (E,2)-1,3-butadiene-1,2,4-tricarboxylic acid by dioxygen
        *note this is the second step in the conversion of

```

```

        4-hydroxybenzoate to succinate and acetyl-CoA
        *superfamily protocatechuate 3,4-dioxygenase alpha chain:
        aromatic hydrocarbon catabolism; dodecamer; heterodimer;
        oxidoreductase
FEATURE
        2-201
        *product protocatechuate 3,4-dioxygenase alpha chain
        *status experimental *label MAL
        *binding_site substrate (Asp) *status predicted
        *length 201 *molecular_weight 22387 *checksum 486
SUMMARY
        134
        Query Match: 43.4% Score 62 DB 1 Length 201
        Best Local Similarity 31.6% Pred. No. 1.06e+01
        Matches 6: Conservative 7: Mismatches 6: Indels 0: Gaps 0:
DB      131  LIFARG:NIHLHTRLYFDEE 149
      1  IFSKLNLIKLNMPLE 19
      1  IFSKLNLIKLNMPLE 19
RESULT  14
ENTRY   *type complete
TITLE   *allergen Amb a 1.1 precursor - common ragweed
ORGANISM *formal_name Ambrosia artemisiifolia *common_name common
        ragweed
DATE    27-Nov-1999 *sequence_revision 03-Apr-1992 *text_change
        29-Sep-1999
ACCESSIONS A39099; A60895; A53240
REFERENCE  A39099
        *authors Rafnar, T.; Griffith, I.J.; Kuc, M.; Bond, J.F.; Rogers,
        B.L.; Klapper, D.G.
        *journal J. Biol. Chem. (1991) 266:1229-1236
        *title Cloning of Amb a 1 (antigen E), the major allergen family of
        short ragweed pollen.
        *cross-references MJD:91093235
        *accession A39099
        *molecule_type mRNA
        *residues 1-396 *label RAF
        *cross-references GB:M63118
REFERENCE  A60895
        *authors Smith, J.J.; Olson, J.R.; Klapper, D.G.
        *journal Mol. Immunol. (1988) 25:355-365
        *title Monoclonal antibodies to denatured ragweed pollen allergen
        Amb a 1: characterization, specificity for the denatured
        allergen, and utilization for the isolation of immunogen-
        peptides of Amb a 1.
        *cross-references MJD:88288254
        *accession A60895
        *molecule_type protein
        *residues 256-273, 292-303, W 305-406 *label SM1
REFERENCE  A53240
        *authors Griffith, I.J.; Pollard, J.; Klapper, D.G.; Rogers, B.L.;
        Nault, A.K.
        *journal Int. Arch. Allergy Appl. Immunol. (1992) 96:296-304
        *title Sequence polymorphism of Amb a 1 and Amb a 11: the major
        allergens in Ambrosia artemisiifolia (short ragweed).
        *cross-references MJD:92234570
        *accession A53240
        *status preliminary
        *molecule_type mRNA
        *residues 1-396 *label GRI
        *cross-references GB:M80559; NID:q166434; PIDN:AAA32565.1; PID:q16435
CLASSIFICATION *superfamily pectate lyase LAF59
KEYWORDS glycoprotein; pol en
FEATURE
        36
        *binding_site carbohydrate (Asn) (covalent) *status
        predicted
SUMMARY
        *length 396 *molecular_weight 42709 *checksum 5506
        Query Match: 43.4% Score 62 DB 2 Length 396
        Best Local Similarity 25.0% Pred. No. 1.06e+01
        Matches 5: Conservative 10: Mismatches 5: Indels 0: Gaps 0:
DB      118  IFEDMWIRLOKEMVNSDK 137

```



.....  
[W][A][T][E][R][M][A][N]  
[P][R][O][T][E][I][N]  
[S][E][Q][U][E][N][C][E]  
[I][N]  
[T][E][R][M]  
.....

Release 3.1A John F. Collins, BioComputing Research Unit.  
Copyright (C) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Mon Jun 19 16:12:55 2000; MasPar time 3.84 seconds  
158,450 Million cell updates/sec

Tabular output not generated.

Title: >US-09-142-524A-11  
Description: (1-20); from US09142524A.ppt  
Perfect Score: 143  
Sequence: 1:FSKNLNKLNMPPLYAGNK 20

Scoring table: FAM 15C  
Gap 15

Searched: 81257 seqs, 30454973 residues

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database: swiss-prot38  
1:swissprot

Statistics: Mean 29.800; Variance 43.129; scale 0.691

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description                        | Pred. No. |
|------------|-------|-------------|--------|----|------------------------------------|-----------|
| 1          | 123   | 79.0        | 374    | 1  | SBP_CRYJA SUGI BASIC PROTEIN PRE   | 2,410-10  |
| 2          | 81    | 55.6        | 327    | 1  | PEL_TORAC PECCIATE LYASE PRECURSO  | 1,558-03  |
| 3          | 74    | 51.7        | 347    | 1  | RH3_YEAST DNA DAMAGE TOLERANCE P   | 2,498-02  |
| 4          | 67    | 46.9        | 392    | 1  | MP14_AMBAR POLLEN ALLERGEN AMB A   | 5,668-01  |
| 5          | 66    | 46.2        | 434    | 1  | PEL_LILLO PECCIATE LYASE PRECURSO  | 1,008-00  |
| 6          | 64    | 44.8        | 395    | 1  | Y154_METJA HYPOTHEICAL PROTEIN M   | 2,278-00  |
| 7          | 63    | 44.1        | 453    | 1  | RBL_RUB11 RIBULOSE BISPHOSPHATE    | 3,318-00  |
| 8          | 63    | 44.1        | 455    | 1  | RBL_LUPM1 RIBULOSE BISPHOSPHATE    | 3,318-00  |
| 9          | 63    | 44.1        | 475    | 1  | RBL_CERGL RIBULOSE BISPHOSPHATE    | 3,318-00  |
| 10         | 63    | 44.1        | 475    | 1  | RBL_STEME RIBULOSE BISPHOSPHATE    | 3,318-00  |
| 11         | 63    | 44.1        | 476    | 1  | RBL_AREDR RIBULOSE BISPHOSPHATE    | 3,318-00  |
| 12         | 63    | 44.1        | 477    | 1  | RBL_AVEDR RIBULOSE BISPHOSPHATE    | 3,318-00  |
| 13         | 62    | 43.4        | 155    | 1  | Y454_MYCGE HYPOTHEICAL PROTEIN M   | 4,898-00  |
| 14         | 62    | 43.4        | 200    | 1  | PCXA_PSEPU PROTOCATECHUATE 3,4-DI  | 4,898-00  |
| 15         | 62    | 43.4        | 389    | 1  | TP6A_SULSH TYPE II DNA TOPOISOMER  | 4,898-00  |
| 16         | 62    | 43.4        | 396    | 1  | MP11_AMBAR POLLEN ALLERGEN AMB A   | 4,898-00  |
| 17         | 62    | 43.4        | 398    | 1  | PE55_LYCES PROBABLE PECCIATE LYASE | 4,898-00  |
| 18         | 62    | 43.4        | 398    | 1  | MP12_AMBAR POLLEN ALLERGEN AMB A   | 4,898-00  |
| 19         | 62    | 43.4        | 1687   | 1  | VIT1_FUNHE VITELLOGENIN II PRECUR  | 4,898-00  |
| 20         | 61    | 42.7        | 115    | 1  | N3M_BALPH NADH:UBIQUINONE OXIDOR   | 7,178-00  |
| 21         | 61    | 42.7        | 393    | 1  | TP6A_METTH TYPE II: DNA TOPO-SOMER | 7,178-00  |
| 22         | 61    | 42.7        | 469    | 1  | RBL_HALR3 RIBULOSE B-SPHOSPHATE    | 7,178-00  |
| 23         | 60    | 42.0        | 253    | 1  | RBL_WAGLA RIBULOSE BISPHOSPHATE    | 1,058-01  |

|    |    |      |     |   |                                    |          |
|----|----|------|-----|---|------------------------------------|----------|
| 24 | 60 | 42.0 | 441 | 1 | BL_CORSA RIBULOSE BISPHOSPHATE     | 1,058-01 |
| 25 | 60 | 42.0 | 443 | 1 | RBL_IRIGE RIBULOSE BISPHOSPHATE    | 1,058-01 |
| 26 | 60 | 42.0 | 443 | 1 | RBL_VERBC RIBULOSE BISPHOSPHATE    | 1,058-01 |
| 27 | 60 | 42.0 | 443 | 1 | RBL_AB_FI RIBULOSE BISPHOSPHATE    | 1,058-01 |
| 28 | 60 | 42.0 | 443 | 1 | RBL_AB_FI RIBULOSE BISPHOSPHATE    | 1,058-01 |
| 29 | 60 | 42.0 | 443 | 1 | RBL_AB_FI RIBULOSE BISPHOSPHATE    | 1,058-01 |
| 30 | 60 | 42.0 | 443 | 1 | RBL_AB_FI RIBULOSE BISPHOSPHATE    | 1,058-01 |
| 31 | 60 | 42.0 | 443 | 1 | RBL_AB_FI RIBULOSE BISPHOSPHATE    | 1,058-01 |
| 32 | 60 | 42.0 | 449 | 1 | PE59_LYCES PROBABLE PECCIATE LYASE | 1,058-01 |
| 33 | 60 | 42.0 | 450 | 1 | RBL_CRAMA RIBULOSE BISPHOSPHATE    | 1,058-01 |
| 34 | 60 | 42.0 | 459 | 1 | RBL_STPLC RIBULOSE BISPHOSPHATE    | 1,058-01 |
| 35 | 60 | 42.0 | 459 | 1 | RBL_NYPER RIBULOSE BISPHOSPHATE    | 1,058-01 |
| 36 | 60 | 42.0 | 465 | 1 | RBL_PASGU RIBULOSE BISPHOSPHATE    | 1,058-01 |
| 37 | 60 | 42.0 | 465 | 1 | RBL_PASGU RIBULOSE BISPHOSPHATE    | 1,058-01 |
| 38 | 60 | 42.0 | 465 | 1 | RBL_PASGU RIBULOSE BISPHOSPHATE    | 1,058-01 |
| 39 | 60 | 42.0 | 467 | 1 | RBL_CALUS RIBULOSE BISPHOSPHATE    | 1,058-01 |
| 40 | 60 | 42.0 | 475 | 1 | RBL_MAGAC RIBULOSE BISPHOSPHATE    | 1,058-01 |
| 41 | 60 | 42.0 | 475 | 1 | RBL_PINRA RIBULOSE BISPHOSPHATE    | 1,058-01 |
| 42 | 60 | 42.0 | 475 | 1 | RBL_PINTH RIBULOSE BISPHOSPHATE    | 1,058-01 |
| 43 | 60 | 42.0 | 475 | 1 | RBL_KETDA RIBULOSE BISPHOSPHATE    | 1,058-01 |
| 44 | 60 | 42.0 | 477 | 1 | RBL_CICIN RIBULOSE BISPHOSPHATE    | 1,058-01 |
| 45 | 60 | 42.0 | 479 | 1 | RBL_BRAOL RIBULOSE BISPHOSPHATE    | 1,058-01 |

ALIGNMENTS

RESULT 1  
ID SBP\_CRYJA STANDARD: PRT: 374 AA.  
AC P18632:  
DI 01-NOV-1990 (Rel. 16, Create  
DI 01-NOV-1995 (Rel. 32, Last se...ence update)  
DE SUGI BASIC PROTEIN PRECURSOR (SBP) (MAJOR ALLERGEN Cry 3.1) (CRY 3.1)  
OS Cryptomeria japonica (Japanese cedar)  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC euphyllophytes; Spermatophytes; Coniferopsida; Coniferales;  
OC Taxodiaceae; Cryptomeria;  
RN [1]  
RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.  
RC TISSUE=POLLEN;  
RX MEDLINE: 94183234  
RA Sore T.; Komiyama N., Shimizu K., Kusakabe T., Morikubo K.,  
RA Kino K.;  
RT Cloning and sequencing of cDNA coding for Cry 3.1, a major allergen  
RT of Japanese cedar pollen.  
RL Biochem. Biophys. Res. Commun. 199;519-525(1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=POLLEN;  
KA Namba M., Karose M., Iorizawa K., Fushida S., Kurimoto M.,  
RN Submitted (JUL-1994) to the EMBL/Genbank/DBJ databases  
RP SEQUENCE OF 22-4...  
RC TISSUE=POLLEN;  
RX MEDLINE: 89031257  
RA Tanai M., Ando S., Usui M., Kurimoto M., Sakaguchi M., Inouye S.,  
RA Matsuhashi T.;  
RT N-terminal amino acid sequence of a major allergen of Japanese cedar  
RT pollen (Cry 3.1).  
RL FEBS Lett. 239;329-332(1988).  
RN [4]  
RP CARBOHYDRATES.  
RX MEDLINE: 95003748.  
RA Hijioka A., Matsumoto I., Kojima K., Ogawa H.;  
RT Antigenicity of the oligosaccharide moiety of the Japanese cedar  
RT (Cryptomeria japonica) pollen allergen, Cry 3.1;  
RL Int. Arch. Allergy Immunol. 105;198-202(1994).  
CC -!- PTN: CONTAINS FUCOSE/XYLOSE-CONTAINING N-LINKED OLIGOSACCHARIDES.  
CC -!- DISEASE: THIS PROTEIN IS THE MAJOR ALLERGEN OF JAPANESE CEDAR  
CC POLLEN. THE MOST COMMON POLLEN ALLERGEN IN JAPAN  
CC -!- MISCELLANEOUS: THE SEQUENCE OF CRY 3.1 FORM A IS SHOWN HERE. FORM  
CC B DIFFERS IN SIX POSITIONS. BELONGS TO THE POLYSACCHARIDE LYASE FAMILY 1.  
CC -!- SIMILARITY: BELONGS TO THE POLYSACCHARIDE LYASE FAMILY 1.

```

CC      AMB A 1/AMB A 11/CRY J 1 SUBFAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL: D26544; BAA05542.1;
CC      EMBL: D26545; BAA05543.1;
CC      EMBL: D34639; BAA0720.1;
CC      PIR: A44773; A44773;
CC      PRAM: PF00544; pec_lyase; 1;
CC      PRINTS: PR00667; AMBALLERGEN;
CC      Allergen: Glycoprotein; Multigene family; Signal.
CC      SIGNAL: 1 21
CC      CHAIN: 22 374
CC      VARIANT 12 12
CC      VARIANT 143 143
CC      VARIANT 222 222
CC      VARIANT 221 221
CC      VARIANT 358 358
CC      VARIANT 361 361
CC      VARIANT 361 361
CC      CARBOHYD 158 158
CC      CARBOHYD 191 191
CC      CARBOHYD 291 291
CC      SEQUENCE 374 AA: 40645 MW: 74AB25950249556F CRC64;
CC      -----
CC      Query Match 79.0%; Score 113; DB 1; Length 374;
CC      Best Local Similarity 75.0%; Pred. No. 2,31e-10;
CC      Matches 15; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
CC      -----
CC      DB 92 1FSKNMKIKLMWYIAGNK 111
CC      QY 1 1FSKNMKIKLMPLYIAGNK 20
CC      -----
CC      RESULT 2
CC      ID PEL_2BAC STANDARD; PRI: 357 AA.
CC      AC Q49721
CC      DT 01-FEB-1995 (Rel. 31; Created)
CC      DT 01-FEB-1995 (Rel. 31; Last sequence update)
CC      DT 15-DEC-1998 (Rel. 37; Last annotation update)
CC      DE PEPTATE LYASE PRECURSOR (EC 4.2.2.2).
CC      OS Nicotiana tabacum (Common tobacco).
CC      OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
CC      euphyllophytes; Spermatophyta; Magnoliophyta; Eudicotyledons;
CC      core eudicots; Asteridae; euasterids I; Solanales; Solanaceae;
CC      Nicotiana.
CC      RN [1]
CC      RP SEQUENCE FROM N.A.
CC      RC STRAIN=CV, SAMSUN; TISSUE=POLLEN;
CC      RX MEDLINE: 93043039.
CC      RA Rogers H.J., Harvey A., Lonsdale D.M.;
CC      RT Isolation and characterization of a tobacco gene with homology to
CC      RT peptate lyase which is specifically expressed during
CC      RT microsporogenesis.
CC      RL Plant Mol. Biol. 20:493-502(1992).
CC      CC 1- CATALYTIC ACTIVITY: ELIMINATIVE CLEAVAGE OF PEPTATE TO GIVE
CC      CC GLYCOSACCHARIDES WITH 4-DEOXY-ALPHA-D-MANN-4-ENURONOSYL GROUPS
CC      CC AT THEIR NON-REDUCING ENDS.
CC      CC 1- DEVELOPMENTAL STAGE: MAXIMALLY EXPRESSED LATE IN POLLEN
CC      CC DEVELOPMENT.
CC      CC 1- SIMILARITY: BELONGS TO THE POLYSACCHARIDE LYASE FAMILY 1.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----

```

```

CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL: X67158; CAA47630.1;
CC      EMBL: X67159; CAA47631.1;
CC      EMBL: X61102; CAA43414.1;
CC      PIR: S26211; S26211;
CC      PRAM: PF00544; pec_lyase; 1;
CC      PRINTS: PR00807; AMBALLERGEN;
CC      Lyase; Signal.
CC      SIGNAL: 1 25
CC      CHAIN: 26 397
CC      ACT_SITE 272 272
CC      FT CARBOHYD 134 134
CC      FT CARBOHYD 227 227
CC      FT CARBOHYD 227 227
CC      FT CONFLICT 156 156
CC      FT CONFLICT 189 189
CC      FT CONFLICT 200 200
CC      FT CONFLICT 202 202
CC      FT CONFLICT 249 249
CC      FT CONFLICT 249 249
CC      SEQUENCE 397 AA: 44351 MW: EF0A22E5A7643F CRC64;
CC      -----
CC      Query Match 56.6%; Score 81; DB 1; Length 397;
CC      Best Local Similarity 50.0%; Pred. No. 1.55e-03;
CC      Matches 10; Conservative 7; Mismatches 3; Indels 0; Gaps 0;
CC      -----
CC      DB 116 1FGKNMKIKLSRELIVTSNK 135
CC      QY 1 1FSKNMKIKLMPLYIAGNK 20
CC      -----
CC      RESULT 3
CC      ID RH31_YEAST STANDARD; PRI: 347 AA.
CC      AC Q05624;
CC      DT 01-NOV-1997 (Rel. 35; Created)
CC      DT 01-NOV-1997 (Rel. 35; Last sequence update)
CC      DT 15-DEC-1998 (Rel. 37; Last annotation update)
CC      DE DNA DAMAGE TOLERANCE PROTEIN P-331 (RAD31 HOMOLOG).
CC      GN AOS1 OR RHC31 OR YPR18CW OR P415.5;
CC      OS Saccharomyces cerevisiae (Baker's yeast).
CC      OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetaceae;
CC      Saccharomyces.
CC      RN [1]
CC      RP SEQUENCE FROM N.A.
CC      RC STRAIN=S288C / AB972;
CC      RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Du Z.,
CC      Favell A., Fulton L., Gattling S., Greco T., Kirstein J.,
CC      Kucaba T., Hallsworth K., Hawkins J., Hillier L., Jier M.,
CC      Johnson D., Johnston L., Landrum V., Lattelle P., Le T.,
CC      Marais E., Meneses S., Miller N., Nham M., Peasey A., Peluse D.,
CC      Rifken L., Riles L., Saito N., Thirumangalakudi V., Vignati G.,
CC      Wilcox C., Wchidman P., Vu H.M., Wilson R., Wierstein K.,
CC      Submitted (MAY-1995) to the EMBL/GenBank/DBJ Databases.
CC      RN [2]
CC      RP CHARACTERIZATION.
CC      RX MEDLINE: 97248694.
CC      RA Shayeghi M., Does C.L., Tavassoli M., Watts F.Z.;
CC      RT Characterization of Schizosaccharomyces pombe rad31, a USA-related
CC      RT gene required for DNA damage tolerance.
CC      RL Nucleic Acids Res. 25:1162-1169(1997).
CC      CC 1- FUNCTION: COULD BE INVOLVED IN A UBIQUITIN-RELATED PROCESS
CC      CC IMPORTANT FOR DNA DAMAGE TOLERANCE.
CC      CC 1- SIMILARITY: TO THE N-TERMINAL OF UBIQUITIN-ACTIVATING ENZYME E1.
CC      CC 1- SIMILARITY: STRONG. TO S.POMBE RAD31.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL: U25842; AAB68113.1;

```



```

DR SEQ: L0004031: AOSL.
DR PFAM: PF003895: ThiF_family; 1.
KW DNA damage.
SC SEQUENCE 347 AA: 39273 MW: 1AD37B817BA27E3E CRC64:

Query Match 51.7% Score 74: DR 1: Length 347:
Best Local Similarity 52.9% Pred. No. 3.49e-02:
Matches 5: Conservative 5: Mismatches 3: Indels 0: Gaps 0:

Dr 140 NTLKRLNIPLYVAGSN 156
      I I I I I I I I I I
QY 4 KNLNKLKMPLYIAGNK 20

RESULT 4
ID MP14_AVRAP STANDARD: PRT: 392 AA.
AC P28749:
D: 01-DEC-1992 (Rel. 24, Created)
D: 01-DEC-1992 (Rel. 24, Last sequence update)
DE POLLEN ALLERGEN AMB A 1.4 PRECURSOR (ANTIGEN 5) (ANTIGEN AMB A 1).
OS Ambrosia artemisiifolia (Short ragweed).
OC Eukaryota: Viridiplantae: Streptophyta: Tracheophyta:
OC euphyllophytes: Spermatophyta: Magnoliophyta: eudicotyledons:
OC core eudicots: Asterales: easterids II: Asterales: Asteraceae:
OC Asteroideae: Heliantheae: Ambrosia.
CC [1]
RN SEQUENCE FROM N.A., AND VARIANTS.
RP TISSUE: POLLEN.
RC MEDLINE: 52234570.
RA Griffith J.J., Pollock J., Klapper D.G., Rogers B.L., Nault A.K.:
"sequence polymorphism of Amb A 1 and Amb A 11, the major allergens
in Ambrosia artemisiifolia (Short ragweed).".
RL Int. Arch. Allergy Appl. Immunol. 96:236-304(1997).
CC -1- SUBUNIT: MONOMER.
CC -1- TISSUE SPECIFICITY: POLLEN AND FLOWERS.
CC -1- DISEASE: ONE OF THE MAJOR ALLERGENS OF THE RAGWEED POLLEN.
CC -1- SIMILARITY: BELONGS TO THE POLYSACCHARIDE LYASE FAMILY 1.
CC AMB A 1/AMB A 11/CRY J 1 SUBFAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@sib-sib.ch).

EMBL: M8C562; AAA32670.1
DR PIR: D53240: D53240.
DR PFAM: PF003844: pec_lyase; 1.
DR PRINTS: PR00807: AMBALLERGEN.
KW Antigen; Allergen; Signal; Multigene family; Polymorphism.
FT SIGNAL 1 25
FT CHAIN 26 392
FT MOD_RES 226 392
FT BLOCKED 182 392
FT VARIANT 182 188 SHGPPV -> CNDGPPA.
SQ SEQUENCE 392 AA: 42842 MW: 7B6219C12F65DA9 CRC64:

Query Match 46.9% Score 67: DB 1: Length 392:
Best Local Similarity 43.0% Pred. No. 6.68e-01:
Matches 7: Conservative 7: Mismatches 6: Indels 0: Gaps 0:

Dr 119 IFARDWIRDLRELAINDK 138
      I I I I I I I I I I
QY 1 IFSKNLNKLKMPLYIAGNK 20

RESULT 5
ID PELLILLO STANDARD: PRT: 434 AA.
AC P40973:
D: 01-FEB-1995 (Rel. 31, Created)
D: 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-DEC-1998 (Rel. 37, Last annotation update)
DE PECTATE LYASE PRECURSOR (EC 4.2.2.2).
OS Lilium longiflorum (Trumpet lily).
OC Eukaryota: Viridiplantae: Streptophyta: Tracheophyta:
OC euphyllophytes: Spermatophyta: Magnoliophyta: Liliopsida: Liliaceae:
OC Liliaceae: Lilium.
CC [1]
RN SEQUENCE FROM N.A.
RP STRAIN: CV. NELLIE WHITE; TISSUE: POLLEN.
RA Kim S.R., Finkel D.J., An G.
RL Submitted (JUN-1993) to the EMBL/GenBank/DBJ databases
CC -1- CATALYTIC ACTIVITY: ELIMINATIVE CLEAVAGE OF PECTATE TO GIVE
CC OLIGOSACCHARIDES WITH 4-DEOXY-ALPHA-D-MANN-4-ENURONIC-5YL GROUPS
CC AT THEIR NON-REDUCING ENDS.
CC -1- SIMILARITY: BELONGS TO THE POLYSACCHARIDE LYASE FAMILY 1.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@sib-sib.ch).

EMBL: Z17328: CAA78976.1
DR EMBL: L18911: AAA33398.1
DR PIR: S29612: S29612.
DR PFAM: PF00544: pec_lyase; 1.
DR PRINTS: PR00807: AMBALLERGEN.
KW Lyase; Signal.
FT SIGNAL 1 22
FT CHAIN 23 434
FT ACT_SITE 312 312
FT POTENTIAL 68
FT CARBOHYD 97 97
FT POTENTIAL 97
SQ SEQUENCE 434 AA: 43457 MW: C1F4E3C4D2BBD364 CRC64:

Query Match 45.2% Score 66: DB 1: Length 434:
Best Local Similarity 43.0% Pred. No. 1.00e+00:
Matches 8: Conservative 6: Mismatches 6: Indels 0: Gaps 0:

Dr 156 IFGKSNVIRLKLQELINNK 175
      I I I I I I I I I I
QY 1 IFSKNLNKLKMPLYIAGNK 20

RESULT 6
ID V144_METIA STANDARD: PRT: 445 AA.
AC G57428:
D: 01-NOV-1997 (Rel. 35, Created)
D: 01-NOV-1997 (Rel. 35, Last sequence update)
D: 01-NOV-1997 (Rel. 35, Last annotation update)
DE HYPO-HEPATIC PROTEIN M30164.
GN M30164.
OS Methanococcus jannaschii.
OC Archaea: Euryarchaeota: Methanococcales: Methanococcaceae:
OC Methanococcus.
CC [1]
RN SEQUENCE FROM N.A.
RP STRAIN: JAL-1 / DSM 2661 / ATCC 43067:
RC MEDLINE: 96337999.
RA Balt C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.:
"Complete genome sequence of the methanogenic archaeon, Methanococcus
jannaschii".
RL Science 273:1056-1073(1996).

```

CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstations.  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).)  
 CC -----  
 DR EMBL: D67473; AAB98152.1;  
 DR XLR: M0284;  
 DR HYPOTHETICAL protein.  
 SQ SEQUENCE 395 AA: 46273 MW: 5E440ABCC9A412 CMC64;

Query Match 44.8% Score 64; DS 1; Length 395;  
 Best Local Similarity 64.3% Pred. No. 2,3e+00;  
 Matches 9; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Db 129 IVEKYLNIKUD:PL 142  
 QY 1 IFSKNLNKLNMP 14

RESULT 7  
 ID RBL\_RUBS1 STANDARD: PHI: 453 AA.  
 AC Q30501  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE RUBULOSE BISPHOSPHATE CARBOXYLASE LARGE CHAIN PRECURSOR (EC 4.1.1.35)  
 CE (RUBISCO LARGE SUBUNIT) (FRAGMENT)  
 GN RBCL  
 OS Rubia tinctorum (Madder).  
 OS Chloroplast.  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;  
 OC core eudicots; Asteridales; euasterids I; Gentianales; Rubiaceae;  
 OC Rubia.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE: 96139325.  
 RA Manc J.F., Nacali A.;  
 RI "Comparison of the evolution of ribulose 1,5-bisphosphate carboxylase  
 RI (rbcL) and atp-rbcL noncoding spacer sequences in a recent plant  
 RI group, the tribe Rubieae (Rubiaceae)."  
 RI J. Mol. Evol. 41:920-927(1995)  
 CC -!- FUNCTION: RUBISCO CATALYZES TWO REACTIONS: THE CARBOXYLATION OF  
 CC D-RIBULOSE 1,5-BISPHOSPHATE, THE PRIMARY EVENT IN PHOTOSYNTHETIC  
 CC CARBON DIOXIDE FIXATION, AS WELL AS THE OXIDATIVE FRAGMENTATION OF  
 CC THE PENTOSE SUBSTRATE IN THE PHOTORESPIRATION PROCESS. BOTH  
 CC REACTIONS OCCUR SIMULTANEOUSLY AND IN COMPETITION AT THE SAME  
 CC ACTIVE SITE.  
 CC -!- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPHOSPHATE - CO(2) -  
 CC 2,3-BISPHOSPHO-D-GLYCERATE.  
 CC -!- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPHOSPHATE + O(2) -  
 CC 3-PHOSPHO-D-GLYCERATE + 2-PHOSPHOGLYCULATE.  
 CC -!- SUBUNIT: 8 LARGE CHAINS + 8 SMALL CHAINS.  
 CC -!- SUBCELLULAR LOCATION: CHLOROPLAST.  
 CC -!- SIMILARITY: BELONGS TO THE RUBISCO LARGE CHAIN FAMILY.  
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).)  
 CC -----  
 DR EMBL: X81104; CAA57010.1;  
 DR HSSP: P06876; RBCL.  
 DR MENDEL: 4673; RUBT1; rbcL1.  
 DR PROSITE: P500157; RUBISCO\_LARGE.1.  
 DR PFAM: PF00016; RuBisCO\_large.1.  
 KW Photosynthesis: Carbon dioxide fixation: Photorespiration;

KW Lyase: Oxidoreductase: Monooxygenase: Chloroplast: Acetylation.  
 FT PROPEP 1 2  
 FT CHAIN 3 >453  
 FT MOD\_RES 3 3  
 FT ACT\_SITE 201 201  
 FT NON\_TER 453 453  
 SQ SEQUENCE 453 AA: 50263 MW: 88653FEA72F849A0 CRC64;

Query Match 44.8% Score 63; DS 1; Length 453;  
 Best Local Similarity 33.3% Pred. No. 3,4e+00;  
 Matches 6; Conservative 7; Mismatches 5; Indels 0;

Db 255 VFARLAVPIVMEYITG 272  
 QY 1 IFSKNLNKLNMP 18

RESULT 8  
 ID RBL\_LUPM1 STANDARD: PHI: 455 AA.  
 AC P92406;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE RUBULOSE BISPHOSPHATE CARBOXYLASE LARGE CHAIN (EC 4.1.1.39) (RUBISCO  
 DE LARGE SUBUNIT) (FRAGMENT).  
 GN RBCL  
 OS Lupinus microcarpus.  
 OS Chloroplast.  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;  
 OC core eudicots; Rosidae; eurosoid I; Fabales; Fabaceae; Papilionoideae;  
 OC Lupinus.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA TISSUE-LEAF.  
 RA Kaess E., Wink M.;  
 RI "Molecular phylogeny of the Papilionaceae (family Leguminosae): first  
 RI sequences versus chemical taxonomy."  
 RI Bot. Acta 108:149-162(1995).  
 CC -!- FUNCTION: RUBISCO CATALYZES TWO REACTIONS: THE CARBOXYLATION OF  
 CC D-RIBULOSE 1,5-BISPHOSPHATE, THE PRIMARY EVENT IN PHOTOSYNTHETIC  
 CC CARBON DIOXIDE FIXATION, AS WELL AS THE OXIDATIVE FRAGMENTATION OF  
 CC THE PENTOSE SUBSTRATE IN THE PHOTORESPIRATION PROCESS. BOTH  
 CC REACTIONS OCCUR SIMULTANEOUSLY AND IN COMPETITION AT THE SAME  
 CC ACTIVE SITE.  
 CC -!- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPHOSPHATE - CO(2) -  
 CC 2,3-BISPHOSPHO-D-GLYCERATE.  
 CC -!- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPHOSPHATE + O(2)  
 CC 3-PHOSPHO-D-GLYCERATE + 2-PHOSPHOGLYCULATE.  
 CC -!- SUBUNIT: 8 LARGE CHAINS + 8 SMALL CHAINS.  
 CC -!- SUBCELLULAR LOCATION: CHLOROPLAST.  
 CC -!- SIMILARITY: BELONGS TO THE RUBISCO LARGE CHAIN FAMILY.  
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).)  
 CC -----  
 DR EMBL: 270063; CAA93922.1;  
 DR PROSITE: P500157; RUBISCO\_LARGE.1.  
 DR PFAM: PF00016; RuBisCO\_large.1.  
 KW Photosynthesis: Carbon dioxide fixation: Photorespiration;  
 KW Lyase: Oxidoreductase: Monooxygenase: Chloroplast.  
 FT NON\_TER 1  
 FT ACT\_SITE 192 192  
 FT NON\_TER 455 455  
 SQ SEQUENCE 455 AA: 50265 MW: 4610CF5DF6145933 CRC64;

Query Match 44.8% Score 63; DS 1; Length 455;

Best Local Similarity 33.3% Pred. No. 3.31e+00; Mismatches 7; Gaps 0;

DB 246 VFARELGVPVIMHDIYT 263  
Matches 6; Conservative 7; Indels 0; Gaps 0;

QY 1 IFSKNLNKLNPLYIAG 18  
Matches 6; Conservative 7; Indels 0; Gaps 0;

RESULT 9

ID RBL\_CERGL STANDARD; PRI: 475 AA.  
AC P2583C;  
DT 01-MAY-1992 (Rel. 22, Created)  
DI 01-MAY-1992 (Rel. 22, Last sequence update)  
DE 15-DEC-1998 (Rel. 37, Last annotation update)  
DE RIBULOSE BISPHOSPHATE CARBOXYLASE LARGE CHAIN PRECURSOR (EC 4.1.1.39)  
DE (RUBISCO LARGE SUBUNIT)  
GN RBCL  
OS Cerastium glomeratum.  
OC Chloroplast.  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;  
OC core eudicots; Caryophyllales; Caryophyllaceae;  
OC Cerastium.  
RN [1]

SEQUENCE FROM N.A.  
RA Manhart J.R., Hugh J.H., Wilson D.;  
RL Submitted (XXX-1992) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: RUBISCO CATALYZES TWO REACTIONS: THE CARBOXYLATION OF  
CC D-RIBULOSE 1,5-BISPHOSPHATE, THE PRIMARY EVENT IN PHOTOSYNTHETIC  
CC CARBON DIOXIDE FIXATION, AS WELL AS THE OXIDATIVE FRAGMENTATION OF  
CC THE PENTOSE SUBSTRATE IN THE PHOTORESPIRATION PROCESS. BOTH  
CC REACTIONS OCCUR SIMULTANEOUSLY AND IN COMPETITION AT THE SAME  
CC ACTIVE SITE.  
CC -!- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPHOSPHATE + CO(2) -  
CC 2 3-PHOSPHO-D-GLYCERATE.  
CC -!- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPHOSPHATE + O(2) -  
CC 3-PHOSPHO-D-GLYCERATE + 2-PHOSPHOGLYCOLATE.  
CC -!- SUBUNIT: 8 LARGE CHAINS + 8 SMALL CHAINS.  
CC -!- SUBCELLULAR LOCATION: CHLOROPLAST.  
CC -!- SIMILARITY: BELONGS TO THE RUBISCO LARGE CHAIN FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: M83542; AAA84173; 1;  
CC RSP: P00875; 1AUS.  
CC MENDEL: 2754; CERGL:rbcl.1.  
CC PROSITE: PS00157; RUBISCO\_LARGE; 1.  
CC PFAM: PF00016; RUBISCO\_LARGE; 1.  
CC Photosynthesis; Carbon dioxide fixation; Photorespiration;  
KW Lyase; Oxidoreductase; Monooxygenase; Chloroplast; Acetylation.  
FT PROPEP 1 2 BY SIMILARITY.  
FT CHAIN 3 475 RIBULOSE BISPHOSPHATE CARBOXYLASE LARGE  
FT CHAIN  
FT MOD\_RES 3 3 ACETYLATION (BY SIMILARITY).  
FT ACT\_SITE 201 201 BINDING OF CO(2) ACTIVATES THE ENZYME.  
SQ SEQUENCE 475 AA: 52656 MW: 60986468E50211B CRC64:

Query Match 44.1% Score 63; DB 1; Length 475;  
Best Local Similarity 33.3% Pred. No. 3.31e+00;  
Matches 6; Conservative 7; Mismatches 5; Indels 0; Gaps 0;

DB 255 VFARELGVPVIMHDIYT 272  
Matches 6; Conservative 7; Indels 0; Gaps 0;

QY 1 IFSKNLNKLNPLYIAG 18  
Matches 6; Conservative 7; Indels 0; Gaps 0;

RESULT 10

ID RBL\_CERGL STANDARD; PRI: 476 AA.  
AC P25827;  
DT 01-MAY-1992 (Rel. 22, Created)  
DI 01-MAY-1992 (Rel. 22, Last sequence update)  
DE 15-DEC-1998 (Rel. 37, Last annotation update)  
DE RIBULOSE BISPHOSPHATE CARBOXYLASE LARGE CHAIN PRECURSOR (EC 4.1.1.39)  
DE (RUBISCO LARGE SUBUNIT)  
GN RBCL  
OS Arenaria drummondii.

Best Local Similarity 33.3% Pred. No. 3.31e+00; Mismatches 7; Gaps 0;

DB 246 VFARELGVPVIMHDIYT 263  
Matches 6; Conservative 7; Indels 0; Gaps 0;

QY 1 IFSKNLNKLNPLYIAG 18  
Matches 6; Conservative 7; Indels 0; Gaps 0;

RESULT 9

ID RBL\_STEME STANDARD; PRI: 475 AA.  
AC P46920;  
DT 01-NOV-1995 (Rel. 32, Created)  
DI 01-NOV-1995 (Rel. 32, Last sequence update)  
DI 15-DEC-1998 (Rel. 37, Last annotation update)  
DE RIBULOSE BISPHOSPHATE CARBOXYLASE LARGE CHAIN PRECURSOR (EC 4.1.1.39)  
DE (RUBISCO LARGE SUBUNIT)  
GN RBCL  
OS Stelleria media.  
OC Chloroplast.  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;  
OC core eudicots; Caryophyllales; Caryophyllaceae;  
OC Stelleria.  
RN [1]

SEQUENCE FROM N.A.  
RA Rettig J.H., Wilson H.D., Manhart J.R.;  
RL Submitted (XXX-1994) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: RUBISCO CATALYZES TWO REACTIONS: THE CARBOXYLATION OF  
CC D-RIBULOSE 1,5-BISPHOSPHATE, THE PRIMARY EVENT IN PHOTOSYNTHETIC  
CC CARBON DIOXIDE FIXATION, AS WELL AS THE OXIDATIVE FRAGMENTATION OF  
CC THE PENTOSE SUBSTRATE IN THE PHOTORESPIRATION PROCESS. BOTH  
CC REACTIONS OCCUR SIMULTANEOUSLY AND IN COMPETITION AT THE SAME  
CC ACTIVE SITE.  
CC -!- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPHOSPHATE + CO(2) -  
CC 2 3-PHOSPHO-D-GLYCERATE.  
CC -!- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPHOSPHATE + O(2) -  
CC 3-PHOSPHO-D-GLYCERATE + 2-PHOSPHOGLYCOLATE.  
CC -!- SUBUNIT: 8 LARGE CHAINS + 3 SMALL CHAINS.  
CC -!- SUBCELLULAR LOCATION: CHLOROPLAST.  
CC -!- SIMILARITY: BELONGS TO THE RUBISCO LARGE CHAIN FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: M62570; AAA84620; 1;  
CC RSP: P00875; 1AUS.  
CC MENDEL: 2993; STEME:rbcl.1.  
CC PROSITE: PS00157; RUBISCO\_LARGE; 1.  
CC PFAM: PF00016; RUBISCO\_LARGE; 1.  
CC Photosynthesis; Carbon dioxide fixation; Photorespiration;  
KW Lyase; Oxidoreductase; Monooxygenase; Chloroplast; Acetylation.  
FT PROPEP 1 2 BY SIMILARITY.  
FT CHAIN 3 475 RIBULOSE BISPHOSPHATE CARBOXYLASE LARGE  
FT CHAIN  
FT MOD\_RES 3 3 ACETYLATION (BY SIMILARITY).  
FT ACT\_SITE 201 201 BINDING OF CO(2) ACTIVATES THE ENZYME.  
SQ SEQUENCE 475 AA: 52675 MW: 5845538C7B585F1 CRC64:

Query Match 44.1% Score 63; DB 1; Length 475;  
Best Local Similarity 33.3% Pred. No. 3.31e+00;  
Matches 6; Conservative 7; Mismatches 5; Indels 0; Gaps 0;

DB 255 VFARELGVPVIMHDIYT 272  
Matches 6; Conservative 7; Indels 0; Gaps 0;

QY 1 IFSKNLNKLNPLYIAG 18  
Matches 6; Conservative 7; Indels 0; Gaps 0;

RESULT 11

ID RBL\_AREDR STANDARD; PRI: 476 AA.  
AC P25827;  
DT 01-MAY-1992 (Rel. 22, Created)  
DI 01-MAY-1992 (Rel. 22, Last sequence update)  
DI 15-DEC-1998 (Rel. 37, Last annotation update)  
DE RIBULOSE BISPHOSPHATE CARBOXYLASE LARGE CHAIN PRECURSOR (EC 4.1.1.39)  
DE (RUBISCO LARGE SUBUNIT)  
GN RBCL  
OS Arenaria drummondii.

CC Chloroplast.  
 CC Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta;  
 CC euphyllophytes: Spermatophyta: Magnoliophyta: eudicotyledons:  
 CC core eudicots: Caryophyllales: Caryophyllaceae: Caryophyllaceae;  
 CC Arenaria.  
 CC [1]  
 RN SEQUENCE FROM N.A.  
 RA Manhart J.R., Huich J.H., Wilson D.J.  
 RC Submitted (XXX-1992) to the EMBL/GenBank/DBJ databases.  
 CC -!- FUNCTION: RUBISCO CATALYZES TWO REACTIONS: THE CARBOXYLATION OF  
 CC D-RIBULOSE 1,5-BISPHOSPHATE, THE PRIMARY EVENT IN PHOTOSYNTHETIC  
 CC CARBON DIOXIDE FIXATION, AS WELL AS THE OXIDATIVE FRAGMENTATION OF  
 CC THE PENTOSE SUBSTRATE IN THE PHOTORESPIRATION PROCESS. BOTH  
 CC REACTIONS OCCUR SIMULTANEOUSLY AND IN COMPETITION AT THE SAME  
 CC ACTIVE SITE.  
 CC -!- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPHOSPHATE + CO(2) =  
 CC 2 3-PHOSPHO-D-GLYCERATE.  
 CC -!- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPHOSPHATE + O(2) =  
 CC 3-PHOSPHO-D-GLYCERATE + 2-PHOSPHOGLYCOLATE.  
 CC -!- SUBUNIT: 8 LARGE CHAINS + 8 SMALL CHAINS.  
 CC -!- SUBCELLULAR LOCATION: CHLOROPLAST.  
 CC -!- SIMILARITY: BELONGS TO THE RUBISCO LARGE CHAIN FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: M83543; AAA84023.1;  
 CC HSP: P08675; LAUS.  
 CC MENDEL: 2732; ARED; rbcL1.  
 CC PROSITE: PS00157; RUBISCO\_LARGE; 1.  
 CC PFAM: PF00016; RuBisCO\_large; 1.  
 CC Photosynthesis: Carbon dioxide fixation: Photorespiration;  
 CC Lyase: Oxidoreductase; Monooxygenase; Chloroplast: Acetylation.  
 CC PROPEP 1 2  
 CC CHAIN 3 476  
 CC RIBULOSE BISPHOSPHATE CARBOXYLASE LARGE  
 CC CHAIN.  
 CC MOD\_RES 3 3  
 CC ACT\_SITE 201 201 BINDING OF CO(2) ACTIVATES THE ENZYME.  
 CC SEQUENCE 476 AA: 52662 MW: 84891304781589 CRO54;  
 CC  
 CC Query Match 44.1%; Score 63; DB 1; Length 476;  
 CC Best Local Similarity 33.3%; Pred. No. 3,31e-00;  
 CC Matches 5; Conservative 7; Mismatches 0; Gaps 0;  
 CC  
 CC DB 255 VFARELGVPIVNHDIYIG 272  
 CC QY 1 IFSKNLNKLNKPLYIAG 18  
 CC  
 CC RESULT 12  
 CC ID RBLAVESA STANDARD: PRT: 477 AA.  
 CC AC P48684;  
 CC DT 01-FEB-1996 (Rel. 33, Created)  
 CC DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 CC DE RIBULOSE BISPHOSPHATE CARBOXYLASE LARGE CHAIN PRECURSOR (EC 4.1.1.39)  
 CC EE (RUBISCO LARGE SUBUNIT).  
 CC GN RBCL.  
 CC OS Avena sativa (Oat).  
 CC OG Chloroplast.  
 CC OC Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta;  
 CC euphyllophytes: Spermatophyta: Magnoliophyta: Liliopsida: Poales;  
 CC Poaceae: Avena.  
 CC [1]  
 CC SEQUENCE FROM N.A.  
 CC Garcia P.G.;  
 CC Submitted (OCT-1993) to the EMBL/GenBank/DBJ databases.  
 CC -!- FUNCTION: RUBISCO CATALYZES TWO REACTIONS: THE CARBOXYLATION OF

CC D-RIBULOSE 1,5-BISPHOSPHATE, THE PRIMARY EVENT IN PHOTOSYNTHETIC  
 CC CARBON DIOXIDE FIXATION, AS WELL AS THE OXIDATIVE FRAGMENTATION OF  
 CC THE PENTOSE SUBSTRATE IN THE PHOTORESPIRATION PROCESS. BOTH  
 CC REACTIONS OCCUR SIMULTANEOUSLY AND IN COMPETITION AT THE SAME  
 CC ACTIVE SITE.  
 CC -!- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPHOSPHATE + CO(2) =  
 CC 2 3-PHOSPHO-D-GLYCERATE.  
 CC -!- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPHOSPHATE + O(2) =  
 CC 3-PHOSPHO-D-GLYCERATE + 2-PHOSPHOGLYCOLATE.  
 CC -!- SUBUNIT: 8 LARGE CHAINS + 8 SMALL CHAINS.  
 CC -!- SUBCELLULAR LOCATION: CHLOROPLAST.  
 CC -!- SIMILARITY: BELONGS TO THE RUBISCO LARGE CHAIN FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: L15300; AAA84028.1;  
 CC HSP: P08676; 3RUB.  
 CC MENDEL: 2735; AVEAR; rbcL1.  
 CC PROSITE: PS00157; RUBISCO\_LARGE; 1.  
 CC PFAM: PF00016; RuBisCO\_large; 1.  
 CC Photosynthesis: Carbon dioxide fixation: Photorespiration;  
 CC Lyase: Oxidoreductase; Monooxygenase; Chloroplast: Acetylation.  
 CC PROPEP 1 2  
 CC CHAIN 3 477  
 CC RIBULOSE BISPHOSPHATE CARBOXYLASE LARGE  
 CC CHAIN.  
 CC MOD\_RES 3 3  
 CC ACT\_SITE 201 201 BINDING OF CO(2) ACTIVATES THE ENZYME.  
 CC SEQUENCE 477 AA: 52935 MW: 84860954505555 CRO54;  
 CC  
 CC Query Match 44.1%; Score 63; DB 1; Length 477;  
 CC Best Local Similarity 33.3%; Pred. No. 3,31e-00;  
 CC Matches 5; Conservative 7; Mismatches 5; Indels 0; Gaps 0;  
 CC  
 CC DB 255 VFARELGVPIVNHDIYIG 272  
 CC QY 1 IFSKNLNKLNKPLYIAG 18  
 CC  
 CC RESULT 13  
 CC ID Y454\_MYCGE STANDARD: PRT: 155 AA.  
 CC AC P47692;  
 CC DT 01-OCT-1996 (Rel. 34, Created)  
 CC DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 CC DT 01-OCT-1996 (Rel. 34, Last annotation update)  
 CC DE HYPOCHLORAL PROTEIN M8454  
 CC GN M8454.  
 CC OS Mycoplasma genitalium.  
 CC OC Bacteria: Firmicutes: Bacillus/Clostridium group: Mollicutes  
 CC Mycoplasmataceae: Mycoplasma.  
 CC [1]  
 CC SEQUENCE FROM N.A.  
 CC STRAIN:ATCC 33530 / G-37;  
 CC MEDLINE: 96026345.  
 CC Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,  
 CC Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,  
 CC Fritchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhrmann J.L.,  
 CC Nguyen D.T., Utterback T.R., Saiki D.K., Phillips C.A., Merrick J.M.,  
 CC Tomb J.-F., Dougherty B.A., Boff K.F., Hu P.-C., Jucier T.S.,  
 CC Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;  
 CC \*The minimal gene complement of Mycoplasma genitalium.\*  
 CC Science 270:397-403(1995).  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial



```

RESULT 15
ID TP6A_SULSH STANDARD: PRI: 389 AA.
AC 005208:
DT 15-JUL-1998 (Rel. 36, Created)
DI 15-JUL-1998 (Rel. 36, Last sequence update)
DI 15-JUL-1998 (Rel. 36, Last annotation update)
DE TYPE 1: DNA TOPOISOMERASE VI SUBUNIT A (EC 5.99.1.3).
GN TOP6A.
OS Sulfolobus shibatae.
OC Archaea: Crenarchaeota: Sulfolobales: Sulfolobus.
RN [1].
RP SEQUENCE FROM N.A., AND SEQUENCE OF 212-221 AND 344-352.
RC STRAIN-B12.
RX MEDLINE: 97238655.
RA Bergerat A., de Massy B., Gadelle D., Voroutas P.-C., Nicolas A.,
RA Forterre P.
RI *An atypical topoisomerase II from Archaea with implications for
RI meiotic recombination.
RL Nature 386:414-417(1997).
CC -- FUNCTION: RELAXES BOTH POSITIVE AND NEGATIVE SUPERURNS AND
CC EXHIBITS A STRONG DECATENASE ACTIVITY.
CC -- CATALYTIC ACTIVITY: ATP-DEPENDENT; BREAKAGE, PASSAGE AND REJOINING
CC OF DOUBLE-STRANDED DNA.
CC -- SUBUNIT: HETEROTETRAMER OF TWO SUBUNITS A AND TWO SUBUNITS B.
CC -- SIMILARITY: BELONGS TO THE TOP6A FAMILY.
CC -----
CC This SWISS-prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Y10582; GAA71635.1;
KW Isomerase; Topoisomerase; DNA-binding.
FT ACT_SITE 106-106 DNA_CLEAVAGE (HY SIMILARITY).
SQ SEQUENCE 389 AA; 45055 MW; 6F617C88231431B7 CRC64;

```

```

Query Match 43.4%; Score 62; DB 1; Length 389;
Best local similarity 46.7%; Pred. No. 4; E-00;
Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

```

```

DB 244 FVRLNEELKLPVYI 258
QY 1 11 11111
QY 2 FSKNLNKLNPVLYI 16

```

```

Search completed: Mon Jun 19 16:13:01 2000
Job time: 5 secs.

```



```

DE POLLEN MAJOR ALLERGEN 1-2.
OS Juniperus ashei (Ozark white cedar).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Coniferopsida; Coniferales;
OC Taxodiaceae; Juniperus.
RN (1)
RP SEQUENCE FROM N.A.
RA MIDORO-HORIUTI I.M., GOLDBLUM R.M., KUROSKY A., WOOD T.G.,
RA BROOKS E.G.
WT *Molecular cloning of mountain cedar (Juniperus ashei) pollen major
WT allergen, Jun a 1.1.
RI Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
CR EMBL: AF106663; AAC03609.1.
CR EMBL: AF106662; AAC03608.1.
CR EMBL: AF106661; AAC03607.1.
CR MENDEL: 36544; Juncus:1388;36544
CR MENDEL: 36545; Juncus:1388;36545
CR MENDEL: 36546; Juncus:1388;36546
SQ SEQUENCE 367 AA; 39824 MW; 402UB693 CF 32;

Query Match 86.0%; Score 123; DB 10; Length 367;
Best Local Similarity 75.0%; Pred. No. 1,470-10;
Matches 15; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 92 IFSQNM:IKKMPLYVAGK 112
QY 1 IFSKNLNKLNMPLYAGK 20
(1) IFSKNLNKLNMPLYAGK 20
(1) IFSKNLNKLNMPLYAGK 20

RESULT 3
ID Q98879 PRELIMINARY: PRT: 444 AA.
AC Q98879
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DI 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DI 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE RIBULOSE BIPHOSPHATE CARBOXYLASE LARGE CHAIN (EC 4.1.1.39)
DE (FRAGMENT).
GN RBCL
OS Conratari tatarici.
OC Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Asterales; Ericales; Lecythidaceae; Conratari.
RN (1)
RP SEQUENCE FROM N.A.
RA MORTON C.M., MORI S.A., PRANCE G.T., KAROL K.G., CHASE M.W.;
RA *Phylogenetic relationships of Lecythidaceae: A cladistic analysis
RA using rbcL sequence and morphological data.
RA Am. J. Bot. 84:530-540(1997).
CC 1- FUNCTION: RIBULOSE CARBOXYLASE TWO REACTIONS: THE CARBOXYLATION OF D-
CC RIBULOSE 1,5-BISPHOSPHATE, THE PRIMARY EVENT IN PHOTOSYNTHESIS;
CC CARBON DIOXIDE FIXATION, AS WELL AS THE OXIDATIVE FRAGMENTATION OF
CC THE PENTOSE SUBSTRATE IN THE PHOTORESPIRATORY PATHWAY. BOTH
CC REACTIONS OCCUR SIMULTANEOUSLY AND IN COMPETITION AT THE SAME
CC ACTIVE SITE.
CC 1- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPHOSPHATE + CO(2) = 2 3-
CC PHOSPHO-D-GLYCERATE.
CC 1- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPHOSPHATE + O(2) = 3-
CC PHOSPHO-D-GLYCERATE + 2-PHOSPHOGLYCOLATE.
CC 1- SUBUNIT: 8 LARGE CHAINS + 8 SMALL CHAINS.
CR EMBL: 230177; CAB0224.1.
CR MENDEL: 36884; Conratari:36884.
CR PROSITE: PS00157; RIBULCO_LARGE.1.
KW Chloroplast; Photosynthesis; Carbon dioxide fixation;
KW Photorespiration; Lyase; Oxidoreductase; Monooxygenase.
FT NONTER 1
FT ACT_SITE 444
FT ACT_SITE 178 178 BINDING OF CO(2) ACTIVATES THE ENZYME.
SQ SEQUENCE 444 AA; 49296 MW; C98E6781 CRC32;

Query Match 47.68; Score 68; DB 8; Length 444;
Best Local Similarity 27.8%; Pred. No. 3,046-00;
Matches 5; Conservative 8; Mismatches 5; Indels 0; Gaps 0;

Db 231 VFARELGPXMHXYLTG 243

```

```

QY 1 IFSKNLNKLNMPLYAG 18
(1) IFSKNLNKLNMPLYAG 18
(1) IFSKNLNKLNMPLYAG 18

RESULT 4
ID Q64510 PRELIMINARY: PRT: 455 AA.
AC Q64510
DI 01-AUG-1998 (TrEMBLrel. 07, Created)
DI 01-AUG-1999 (TrEMBLrel. 07, Last sequence update)
DI 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE PUTATIVE PEPTIDE HYDROLASE.
DE 120F6.14.
GN Arabidopsis thaliana (Mouse-ear cress);
OS Arabidopsis thaliana; Streptophyta; Embryophyta; Tracheophyta;
OC Eukaryota; Viridiplantae; Streptophyta; Magnoliophyta; eudicotyledons;
OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Rosidae; eurosids I; Brassicales; Brassicaceae;
OC Arabidopsis.
RN (1)
RP SEQUENCE FROM N.A.
RA STRAIN-OLY. COLUMBIA;
RA ROUNSELEY S.D., LIN X., KETCHUM K.A., GRUSHY M.L., BRAND N.W.;
RA SYKES S.M., KAUL S., WASON T.M., KIELAVAGE A.R., ADAMS M.C.;
RA SOMERVILLE C.R., VENTER J.C.;
RA Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
CR EMBL: AAC02521; AAC05950.1.
CR MENDEL: 28505; Arabid:1088;28505.
CR PFAM: PF00544; pec_lyase.1.
CR PRINTS: PRC0807; AMBALLERGEN.
KW Lyase
SQ SEQUENCE 455 AA; 51257 MW; E5ACH62D CRC32;

Query Match 47.68; Score 68; DB 10; Length 455;
Best Local Similarity 40.0%; Pred. No. 3,046-00;
Matches 8; Conservative 6; Mismatches 6; Indels 0; Gaps 0;

Db 171 IFSMIKLOQELIITNK 19
QY 1 IFSKNLNKLNMPLYAGK 20
(1) IFSKNLNKLNMPLYAGK 20
(1) IFSKNLNKLNMPLYAGK 20

RESULT 5
ID Q21303 PRELIMINARY: PRT: 1202 AA.
AC Q21303
DI 01-NOV-1996 (TrEMBLrel. 01, Created)
DI 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DI 01-JAN-1999 (TrEMBLrel. 09, Last annotation update)
DE K07G5.3 PROTEIN.
GN K07G5.3
OS Coccidioides immitis.
OC Eukaryota; Metazoa; Nematoda; Streptophyta; Rhodophyta; Rhodophyta;
OC Rhodophyta; Rhodophyta; Rhodophyta; Rhodophyta; Rhodophyta; Rhodophyta;
RN (1)
RP SEQUENCE FROM N.A.
RA MCMURRAY A.;
RA Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
RN (2)
RP SEQUENCE FROM N.A.
RA MEDLINE: 94150719.
RA WILSON R., AINSWORTH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COLLISON A.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KIRSTEN N., LAISTER N., LAKEILJE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIFKIN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SMAILDON N., SMITH A., SONNHAMMER E., STADEN R., SULTON J.,
RA THIERVY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOLDMAN P.;
RT *2.2 Mb of contiguous nucleotide sequence from chromosome 11; of 2.
RT elegans.
RL Nature 368:32-38(1994).
DR EMBL: 271264; GAA95830.1.
SQ SEQUENCE 1202 AA; 138459 MW; 6F9C9E53 CRC32;

```



Query Match: 47.6%; Score 68; DB 5; Length 1202;  
 Best Local Similarity 51.5%; Pred. No. 3,04e+00;  
 Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 8:1 IRLNPLYISNYK 823  
 QY 8 IRLNPLYISNYK 20

RESULT 6  
 ID 078617 PRELIMINARY: PRT: 430 AA.  
 AC 078617  
 DT 01-NOV-1998 (TREMBlrel. 08, Created)  
 DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)  
 DT 01-NOV-1998 (TREMBlrel. 12, Last annotation update)  
 DE RIBULOSE BIPHOSPHATE CARBOXYLASE LARGE CHAIN (EC 4.1.1.39)  
 DE (FRAGMENT);  
 GN RBCL;  
 OS Chlorococcum trnasutum;  
 OS Chloroplast;  
 OS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OS eubryophytes; Spermatophyta; Magnoliophyta; Liliopsida; Asparagales;  
 OS Orchidaceae; Meliracillium;  
 RN SEQUENCE FROM N.A.  
 RA CAMERON K.M., CHASE M.W., WHITTEN W.M., KORES P.J., JARRELL D.C.,  
 RA ALBERT V.A., YUKAWA T., HILLS H.G., GOLDMAN D.H.;  
 RT "A phylogenetic analysis of the Orchidaceae: evidence from rbcL  
 nucleotide sequences";  
 RC Am. J. Bot. 0:0-0(1998);

CC -!- FUNCTION: RUBISCO CATALYZES TWO REACTIONS: THE CARBOXYLATION OF D-  
 RIBULOSE 1,5-BISPHOSPHATE, THE PRIMARY EVENT IN PHOTOSYNTHETIC  
 CARBON DIOXIDE FIXATION, AS WELL AS THE OXIDATIVE FRAGMENTATION OF  
 THE PENTOSE SUBSTRATE IN THE PHOTORESPIRATION PROCESS. BOTH  
 REACTIONS OCCUR SIMULTANEOUSLY AND IN COMPETITION AT THE SAME  
 ACTIVE SITE.  
 CC -!- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPHOSPHATE + CO(2) -> 2 3-  
 PHOSPHO-D-GLYCERATE;  
 CC -!- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPHOSPHATE + O(2) -> 3-  
 PHOSPHO-D-GLYCERATE + 2-PHOSPHOGLYCOLATE;  
 CC -!- SUBUNIT: 8 LARGE CHAINS + 8 SMALL CHAINS.  
 CC EMBL: AF074192; AAC35168.1;  
 DR HSP: P04718; GRUB.  
 DR MENDEL: 30551; Meltr:rbcl30551;  
 DR PROSITE: PS00157; RUBISCO\_LARGE; 1.  
 DR PFAM: PF00016; RUBISCO\_LARGE; 1.  
 KW Chloroplast; Photosynthesis; Carbon dioxide fixation;  
 KW Photorespiration; Lyase; Oxidoreductase; Monooxygenase.  
 FT NON-TER;  
 FT ACT\_SITE 182 430 BINDING OF CO(2) ACTIVATES THE ENZYME.  
 FT ACT\_SITE 182 182  
 SQ SEQUENCE 430 AA: 47681 MW: 759E959E CRC32:

Query Match: 45.9%; Score 57; DB 8; Length 430;  
 Best Local Similarity 33.3%; Pred. No. 4.39e+00;  
 Matches 6; Conservative 8; Mismatches 4; Indels 0; Gaps 0;

Db 235 VFARELGVFIMPDYLTG 252  
 QY 1 IFSKNLNKLNPLYIAGK 18

RESULT 7  
 ID 023665 PRELIMINARY: PRT: 459 AA.  
 AC 023665;  
 DT 01-JAN-1998 (TREMBlrel. 05, Created)  
 DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)  
 DT 01-NOV-1998 (TREMBlrel. 12, Last annotation update)  
 DE PUTATIVE PECTATE LYASE;  
 GN 59;  
 OS Arabidopsis thaliana (Mouse-ear cress);  
 OS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;  
 OC core eudicots; Rosidae; eurosids 1; Brassicales; Brassicaceae;  
 OC Arabidopsis;  
 RN SEQUENCE FROM N.A.  
 RP MEDLINE: 97422403.  
 RA KULIKASKAS R., MCCORMICK S.;  
 RT "Identification of the tobacco and Arabidopsis homologues of the  
 RI pollen-expressed LAT59 gene of tomato";  
 RL Plant Mol. Biol. 34:809-814(1997).  
 DR EMBL: U03619; AAB69759.1;  
 DR MENDEL: 25559; Arabid088; 25559.  
 DR PFAM: PF00544; pect\_lyase; 1.  
 DR PRINTS: PRO0807; AMBALLERGEN;  
 KW Lyase.  
 SQ SEQUENCE 459 AA: 51420 MW: 41EBE908 CRC32:

Query Match: 46.9%; Score 67; DB 10; Length 459;  
 Best Local Similarity 40.0%; Pred. No. 4.39e+00;  
 Matches 8; Conservative 7; Mismatches 5; Indels 0; Gaps 0;

Db 175 IFARSMIKLQQLMITSXK 194  
 QY 1 IFSKNLNKLNPLYIAGK 20

RESULT 8  
 ID 018771 PRELIMINARY: PRT: 555 AA.  
 AC 018771;  
 DT 01-NOV-1996 (TREMBlrel. 01, Created)  
 DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)  
 DT 01-MAY-1999 (TREMBlrel. 10, Last annotation update)  
 DE CS1E3.6 PROTEIN.  
 GN CS1E3.6;  
 OS Caenorhabditis elegans.  
 OS Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;  
 OC Rhabditina; Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis;  
 RN SEQUENCE FROM N.A.  
 RA WILKINSON J.;  
 RL Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.  
 RN SEQUENCE FROM N.A.  
 RP MEDLINE: 94150718.  
 RA WILSON R., AINSOUGH R., ANDERSON K., BAYNES C., BERKS M.,  
 RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COLLINS N.A.,  
 RA CRAXTON M., DEAR S., DU Z., DUBIN R., FAVELLO A., FULTON L.,  
 RA GARDNER A., GREEN P., HANKINS J., HILLIER L., JIER M., JOHNSTON L.,  
 RA JONES M., KERSHAW J., KIRSTEN J., LAISER N., LAURENCE P.,  
 RA LIGHTNING J., LLOYD C., MENDRIPAY A., MORTIMER R., O'CALLAGHAN M.,  
 RA PARSONS J., PERRY C., RIFKEN L., POLPRA A., SAUNDERS D., SHANKEN P.,  
 RA SKALDON N., SMITH A., SWANHAMMER E., STADEN R., SULTON J.,  
 RA THIERRY-MIEG J., THOMAS K., VANDIN M., VAUGHAN K., WATERSTON K.,  
 RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT C., WOHLMAN P.,  
 RT "2 Mb of contiguous nucleotide sequence from chromosome III of C.  
 elegans";  
 RL Nature 368:32-38(1994).  
 DR EMBL: 278410; CAB01641.1;  
 DR PFAM: PF00860; xan\_ur\_permease; 1.  
 SQ SEQUENCE 555 AA: 60315 MW: 4D90DD42 CRC32:

Query Match: 46.9%; Score 67; DB 5; Length 555;  
 Best Local Similarity 35.3%; Pred. No. 4.39e+00;  
 Matches 6; Conservative 7; Mismatches 4; Indels 0; Gaps 0;

Db 464 FNOILNVLLQMPFVGA 480  
 QY 2 FSKNLNKLNPLYIAG 18

RESULT 9  
 ID 031994 PRELIMINARY: PRT: 414 AA.  
 AC 031994;

01-NOV-1996 (TRENBLREI. 01, Created)  
 01-NOV-1996 (TRENBLREI. 01, Last sequence update)  
 01-NOV-1999 (TRENBLREI. 12, Last annotation update)  
 RIBULOSE BISPSPHOSPHATE CARBOXYLASE LARGE CHAIN (EC 4.1.1.39)  
 (FRAGMENT).  
 DB 240 IFARELGVPVIMHDYITG 257  
 QY 1 IFSKNINIKLNPLYIAG 18  
 RESULT 10 PRELIMINARY: PRT: 445 AA.  
 AC Q33403:  
 DT 01-NOV-1996 (TRENBLREI. 01, Created)  
 DT 01-NOV-1996 (TRENBLREI. 01, Last sequence update)  
 DT 01-NOV-1999 (TRENBLREI. 12, Last annotation update)  
 DE RIBULOSE BISPSPHOSPHATE CARBOXYLASE LARGE CHAIN (EC 4.1.1.39)  
 (FRAGMENT).  
 CE RECL.  
 CS Donatia fascicularis.  
 CG Chloroplast.  
 CC Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta:  
 euphyllophytes: Spermatophyta: Magnoliophyta: eudicotyledons:  
 core eudicots: Asteridae: euasterids I: Asterales: Donatiaceae:  
 Donatia.  
 RN [1]  
 FP SEQUENCE FROM N.A.  
 PA GUSTAFSSON M.H.G., BACKLUND A., BREMER B.:  
 Plant Syst. Evol. 199:217-242(1995).  
 RL -1- FUNCTION: RUBISCO CATALYSES TWO REACTIONS: THE CARBOXYLATION OF D-  
 RIBULOSE 1,5-BISPSPHOSPHATE, THE PRIMARY EVENT IN PHOTOSYNTHETIC  
 CARBON DIOXIDE FIXATION, AS WELL AS THE OXIDATIVE FRAGMENTATION OF  
 THE PENTOSE SUBSTRATE IN THE PHOTORESPIRATION PROCESS. BOTH  
 REACTIONS OCCUR SIMULTANEOUSLY AND IN COMPETITION AT THE SAME  
 ACTIVE SITE.  
 CC -1- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPSPHOSPHATE + CO(2) -> 2 3-  
 PHOSPHO-D-GLYCERATE.  
 CC -1- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPSPHOSPHATE + O(2) -> 3-  
 PHOSPHO-D-GLYCERATE + 2-PHOSPHOGLYCOLATE.  
 CC -1- SUBUNIT: 8 LARGE CHAINS - 8 SMALL CHAINS.  
 CC EMBL: D44563; RA07967.1; .  
 DR MENDEL: 21634; Coqem:rbcl:21634.  
 DR PROSITE: PSC0157; RUBISCO\_LARGE: 1.  
 DR PFAM: PF00016; RUBISCO\_LARGE: 1.  
 DR Chloroplast: Photosynthesis: Carbon dioxide fixation:  
 Photorespiration: Lyase: Oxidoreductase; Monooxygenase.  
 FN NON-TER 414 414  
 FT ACT\_SITE 187 187 BINDING OF CO(2) ACTIVATES THE ENZYME.  
 SQ SEQUENCE 414 AA: 45839 MW: 581502 CRC32:  
 Query Match 45.5% Score 65: DB 8: Length 414;  
 Best Local Similarity 38.9% Pred. No. 9.08e+00;  
 Matches 7: Conservative 6: Mismatches 5: Indels 0: Gaps 0:

CC REACTIONS OCCUR SIMULTANEOUSLY AND IN COMPETITION AT THE SAME  
 CC ACTIVE SITE.  
 CC -1- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPSPHOSPHATE + CO(2) -> 2 3-  
 PHOSPHO-D-GLYCERATE.  
 CC -1- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPSPHOSPHATE + O(2) -> 3-  
 PHOSPHO-D-GLYCERATE + 2-PHOSPHOGLYCOLATE.  
 CC -1- SUBUNIT: 8 LARGE CHAINS - 8 SMALL CHAINS.  
 CC EMBL: X87385; CA60839.1; .  
 DR MENDEL: 21722; Donatia:rbcl:21722.  
 DR PROSITE: PSC0157; RUBISCO\_LARGE: 1.  
 DR PFAM: PF00016; RUBISCO\_LARGE: 1.  
 DR Chloroplast: Photosynthesis: Carbon dioxide fixation:  
 Photorespiration: Lyase: Oxidoreductase; Monooxygenase.  
 FN NON-TER 193 193  
 FT ACT\_SITE 193 193 BINDING OF CO(2) ACTIVATES THE ENZYME.  
 SQ SEQUENCE 449 AA: 45527 MW: 5857102 CRC32:  
 Query Match 45.5% Score 65: DB 8: Length 449;  
 Best Local Similarity 38.9% Pred. No. 9.08e+00;  
 Matches 7: Conservative 6: Mismatches 5: Indels 0: Gaps 0:

DB 246 IFARELGVPVIMHDYITG 263  
 QY 1 IFSKNINIKLNPLYIAG 18  
 RESULT 11 PRELIMINARY: PRT: 456 AA.  
 ID Q9XCF1:  
 AC Q9XCF1:  
 DT 01-NOV-1999 (TRENBLREI. 12, Created)  
 DT 01-NOV-1999 (TRENBLREI. 12, Last sequence update)  
 DT 01-NOV-1999 (TRENBLREI. 12, Last annotation update)  
 DE RIBULOSE BISPSPHOSPHATE CARBOXYLASE LARGE CHAIN (EC 4.1.1.39)  
 (FRAGMENT).  
 CE RECL.  
 CS Donatia novae-zelandiae.  
 CG Chloroplast.  
 CC Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta:  
 euphyllophytes: Spermatophyta: Magnoliophyta: eudicotyledons: core  
 eudicots: Asteridae: euasterids I: Asterales: Donatiaceae: Donatia.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-LEAF:  
 RA LAURENT N.C., BREMER B., BREMER K.:  
 "Phylogeny and genetic interrelationships of the Stadiaceae  
 (Asteridae) with a possible outgroup from the floral pseudomorphaceae".  
 Submitted (Mar-1998) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: RUBISCO CATALYSES TWO REACTIONS: THE CARBOXYLATION OF D-  
 RIBULOSE 1,5-BISPSPHOSPHATE, THE PRIMARY EVENT IN PHOTOSYNTHETIC  
 CARBON DIOXIDE FIXATION, AS WELL AS THE OXIDATIVE FRAGMENTATION OF  
 THE PENTOSE SUBSTRATE IN THE PHOTORESPIRATION PROCESS. BOTH  
 REACTIONS OCCUR SIMULTANEOUSLY AND IN COMPETITION AT THE SAME  
 ACTIVE SITE.  
 CC -1- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPSPHOSPHATE + CO(2) -> 2 3-  
 PHOSPHO-D-GLYCERATE.  
 CC -1- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPSPHOSPHATE + O(2) -> 3-  
 PHOSPHO-D-GLYCERATE + 2-PHOSPHOGLYCOLATE.  
 CC -1- SUBUNIT: 8 LARGE CHAINS - 8 SMALL CHAINS.  
 CC EMBL: A225058; CAB39389.1; .  
 DR PROSITE: PSC0157; RUBISCO\_LARGE: 1.  
 DR Chloroplast: Photosynthesis: Carbon dioxide fixation:  
 Photorespiration: Lyase: Oxidoreductase; Monooxygenase.  
 FN NON-TER 193 193  
 FT ACT\_SITE 193 193 BINDING OF CO(2) ACTIVATES THE ENZYME.  
 SQ SEQUENCE 466 AA: 51570 MW: 504849A CRC32:  
 Query Match 45.5% Score 65: DB 8: Length 466;  
 Best Local Similarity 38.9% Pred. No. 9.08e+00;  
 Matches 7: Conservative 6: Mismatches 5: Indels 0: Gaps 0:

Db 246 IFARELGVPIVMHDIYIG 263  
 LU : : : : :  
 QY 1 IFSKNLKLNPLYIAG 18  
 : : : : :  
 RESULT 12  
 AC P93942 PRELIMINARY: PRT: 467 AA.  
 CT 01-MAY-1997 (Tremblrel. 03, Created)  
 DT 01-MAY-1997 (Tremblrel. 03, Last sequence update)  
 DT 01-NOV-1999 (Tremblrel. 12, Last annotation update)  
 DE RIBULOSE BISPHOSPHATE CARBOXYLASE LARGE CHAIN (EC 4.1.1.39)  
 CE (FRAGMENT).  
 GN RBCL.  
 OS Odontostomum hartwegii.  
 OC Chloroplast.  
 OC Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta:  
 OC euphyllophytes: Spermatophyta: Magnoliophyta: Liliopsida: Asparagales:  
 OC Tecophilaeaceae: Odontostomum.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-LEAF:  
 RA RUDALL P.J., FURNESS C.A., FAY M.F., CHASE M.W.:  
 RL Submitted (JUL1996) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: RUBISCO CATALYZES TWO REACTIONS: THE CARBOXYLATION OF D-  
 CC RIBULOSE 1,5-BISPHOSPHATE: THE PRIMARY EVENT IN PHOTOSYNTHETIC  
 CC CARBON DIOXIDE FIXATION, AS WELL AS THE OXIDATIVE FRAGMENTATION OF  
 CC THE PENTOSE SUBSTRATE IN THE PHOTORESPIRATION PROCESS. BOTH  
 CC REACTIONS OCCUR SIMULTANEOUSLY AND IN COMPETITION AT THE SAME  
 CC ACTIVE SITE.  
 CC -1- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPHOSPHATE + CO(2) -> 2 3-  
 CC PHOSPHO-D-GLYCERATE.  
 CC -1- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPHOSPHATE + O(2) -> 3-  
 CC PHOSPHO-D-GLYCERATE + 2-PHOSPHOGLYCOLATE.  
 CC -1- SUBUNIT: 8 LARGE CHAINS + 8 SMALL CHAINS.  
 CC -1- SUBCELLULAR LOCATION: CHLOROPLAST.  
 CC EMBL: Z77314; CAB01109.1;  
 DR HSP: P06830; IRSC.  
 DR MENDEL: 14408; OdoA:rbcl.14408.  
 DR PROSITE: PS00157; RUBISCO\_LARGE.1.  
 DR PFAM: PF00016; RUBISCO\_LARGE.1.  
 DR Lyase: Chloroplast; Photosynthesis: Carbon dioxide fixation;  
 KW Photorespiration: Oxidoreductase; Monooxygenase.  
 FT NON\_TER 1  
 FT ACT\_SITE 192 192 BINDING OF CO(2) ACTIVATES THE ENZYME.  
 FT ACT\_SITE 193 193 BINDING OF CO(2) ACTIVATES THE ENZYME.  
 FT NON\_TER 467 467  
 SQ SEQUENCE 467 AA: 51035 MW: 1618A356 CRC32:  
 Query Match 45.5% Score 65: DB 8: Length 467:  
 Best Local Similarity 27.8% Pred. No. 9,08e+00:  
 Matches 5: Conservative 8: Mismatches 5: Indels 0: Gaps 0:  
 Db 246 VFARELGVPIVMHXYLIIG 263  
 LU : : : : :  
 QY 1 IFSKNLKLNPLYIAG 18  
 : : : : :  
 RESULT 13  
 AC Q31807 PRELIMINARY: PRT: 475 AA.  
 CT 01-NOV-1996 (Tremblrel. 01, Created)  
 DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)  
 DT 01-NOV-1999 (Tremblrel. 12, Last annotation update)  
 DE RIBULOSE BISPHOSPHATE CARBOXYLASE LARGE CHAIN (EC 4.1.1.39)  
 CE (FRAGMENT).  
 GN RBCL.  
 OS Angelica lucida.  
 OC Chloroplast.  
 OC Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta:  
 OC euphyllophytes: Spermatophyta: Magnoliophyta: eudicotyledons:  
 OC core eudicots: Asteridae: euasterids II: Apiales: Apiaceae: Angelica.  
 RN [1]

RP SEQUENCE FROM N.A.  
 RA PLUNKETT G.M., SOLTIS D.E., SOLTIS P.S.:  
 RL Am. J. Bot. 83:499-515(1996).  
 CC -1- FUNCTION: RUBISCO CATALYZES TWO REACTIONS: THE CARBOXYLATION OF D-  
 CC RIBULOSE 1,5-BISPHOSPHATE: THE PRIMARY EVENT IN PHOTOSYNTHETIC  
 CC CARBON DIOXIDE FIXATION, AS WELL AS THE OXIDATIVE FRAGMENTATION OF  
 CC THE PENTOSE SUBSTRATE IN THE PHOTORESPIRATION PROCESS. BOTH  
 CC REACTIONS OCCUR SIMULTANEOUSLY AND IN COMPETITION AT THE SAME  
 CC ACTIVE SITE.  
 CC -1- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPHOSPHATE + CO(2) -> 2 3-  
 CC PHOSPHO-D-GLYCERATE.  
 CC -1- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPHOSPHATE + O(2) -> 3-  
 CC PHOSPHO-D-GLYCERATE + 2-PHOSPHOGLYCOLATE.  
 CC -1- SUBUNIT: 8 LARGE CHAINS + 8 SMALL CHAINS.  
 CC -1- SUBCELLULAR LOCATION: CHLOROPLAST.  
 CC EMBL: U50221; AAA98935.1;  
 DR MENDEL: 21616; Anglu:rbcl.21616.  
 DR PROSITE: PS00157; RUBISCO\_LARGE.1.  
 DR PFAM: PF00016; RUBISCO\_LARGE.1.  
 DR Chloroplast; Photosynthesis: Carbon dioxide fixation;  
 KW Photorespiration: Lyase: Oxidoreductase; Monooxygenase.  
 FT NON\_TER 1  
 FT ACT\_SITE 201 201 BINDING OF CO(2) ACTIVATES THE ENZYME.  
 FT ACT\_SITE 202 202 BINDING OF CO(2) ACTIVATES THE ENZYME.  
 SQ SEQUENCE 475 AA: 52676 MW: BE991079 CRC32:  
 Query Match 45.5% Score 65: DB 8: Length 475:  
 Best Local Similarity 36.9% Pred. No. 9,08e+00:  
 Matches 7: Conservative 6: Mismatches 5: Indels 0: Gaps 0:  
 Db 255 IFARELGVPIVMHDIYIG 272  
 LU : : : : :  
 QY 1 IFSKNLKLNPLYIAG 18  
 : : : : :  
 RESULT 14  
 ID P92302 PRELIMINARY: PRT: 475 AA.  
 AC P92302:  
 DT 01-MAY-1997 (Tremblrel. 03, Created)  
 DT 01-MAY-1997 (Tremblrel. 03, Last sequence update)  
 DT 01-NOV-1999 (Tremblrel. 12, Last annotation update)  
 DE RIBULOSE BISPHOSPHATE CARBOXYLASE LARGE CHAIN (EC 4.1.1.39).  
 GN RBCL.  
 OS Hillebrandia sandwicensis.  
 OG Chloroplast.  
 OC Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta:  
 OC euphyllophytes: Spermatophyta: Magnoliophyta: eudicotyledons:  
 OC core eudicots: Rosidae: euros II: Cucurbitales: Begoniaceae.  
 OC Hillebrandia.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA SWENSEN S.M.:  
 RL Am. J. Bot. 83:1503-1512(1996).  
 CC -1- FUNCTION: RUBISCO CATALYZES TWO REACTIONS: THE CARBOXYLATION OF D-  
 CC RIBULOSE 1,5-BISPHOSPHATE: THE PRIMARY EVENT IN PHOTOSYNTHETIC  
 CC CARBON DIOXIDE FIXATION, AS WELL AS THE OXIDATIVE FRAGMENTATION OF  
 CC THE PENTOSE SUBSTRATE IN THE PHOTORESPIRATION PROCESS. BOTH  
 CC REACTIONS OCCUR SIMULTANEOUSLY AND IN COMPETITION AT THE SAME  
 CC ACTIVE SITE.  
 CC -1- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPHOSPHATE + CO(2) -> 2 3-  
 CC PHOSPHO-D-GLYCERATE.  
 CC -1- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPHOSPHATE + O(2) -> 3-  
 CC PHOSPHO-D-GLYCERATE + 2-PHOSPHOGLYCOLATE.  
 CC -1- SUBUNIT: 8 LARGE CHAINS + 8 SMALL CHAINS.  
 CC -1- SUBCELLULAR LOCATION: CHLOROPLAST.  
 CC EMBL: U59822; AAB41157.1;  
 DR HSP: P00875; IRCL.  
 DR MENDEL: 21804; Hilsa:rbcl.21804.  
 DR PROSITE: PS00157; RUBISCO\_LARGE.1.  
 DR PFAM: PF00016; RUBISCO\_LARGE.1.  
 DR Chloroplast; Photosynthesis: Carbon dioxide fixation;  
 KW Photorespiration: Lyase: Oxidoreductase; Monooxygenase.  
 FT ACT\_SITE 201 201 BINDING OF CO(2) ACTIVATES THE ENZYME.

FT ACT\_SITE 202 202 BINDING OF CO(2) ACTIVATES THE ENZYME.  
SQ SEQUENCE 475 AA: 52775 MW: DC4CA579 CR032.

Query Match 45.5% Score 65: DB 8: Length 475:  
Best Local Similarity 38.9% Pred. NO. 9,08e-00:  
Matches 7: Conservative 6: Mismatches 5: Indels 0: Gaps 0:

CU 255 :FARELGVPIVNHDIITG 272  
:|||||:|:|:  
CV 1 :FSKNLNKLNMPYIAG 18

RESULT 15  
ID OC9485 PRELIMINARY: PRT: 475 AA  
AC OC9485  
D: 01-JUL-1997 (TrEMBLrel\_04: Created)  
D: 01-JUL-1997 (TrEMBLrel\_04: Last sequence update)  
DI 01-NOV-1999 (TrEMBLrel\_12: Last annotation update)  
DE RIBULOSE BISPHOSPHATE CARBOXYLASE LARGE CHAIN (EC 4.1.1.39)  
DE (FRAGMENT)  
GN RBCL  
OS Hydrostachys sp. 'Bremer 3089'  
OG Chloroplast  
OC Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta:  
OC euphyllophytes: Spermatophyta: Magnoliophyta: eudicotyledons:  
OC core eudicots: Asteridae: Cornales: Hydrostachyaceae: Hydrostachys  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-COLL: BREMER 3089 (UPS)  
RA BACKLUND A., BREMER B.  
RL Plant Syst. Evol. 200:3-6(1997).  
CC -!- FUNCTION: RUBISCO CATALYSES TWO REACTIONS: THE CARBOXYLATION OF D-  
RIBULOSE 1,5-BISPHOSPHATE, THE PRIMARY EVENT IN PHOTOSYNTHETIC  
CARBON DIOXIDE FIXATION, AS WELL AS THE OXIDATIVE FRAGMENTATION OF  
THE PENTOSE SUBSTRATE IN THE PHOTORESPIRATION PROCESS. BOTH  
REACTIONS OCCUR SIMULTANEOUSLY AND IN COMPETITION AT THE SAME  
ACTIVE SITE.  
CC -!- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPHOSPHATE + CO(2) = 2 3-  
PHOSPHO-D-GLYCERATE.  
CC -!- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPHOSPHATE + O(2) = 3-  
PHOSPHO-D-GLYCERATE + 2-PROSPHOGLYCOLATE.  
CC -!- SUBUNIT: 8 LARGE CHAINS + 8 SMALL CHAINS.  
CC -!- SUBCELLULAR LOCATION: CHLOROPLAST.  
DR EMBL: Y10708; CAA71713.1;  
DR MENDEL: 21881; Hydsp:rbcc21881.  
DR PROSITE: PS00157; RUBISCOLARGE: 1.  
DR PFAM: PF00616; RUBISCOLARGE: 1.  
KW Lyase: Chloroplast; Photosynthesis; Carbon dioxide fixation;  
KW Photorespiration; Oxidoreductase; Monooxygenase.  
FT NON\_TER 1  
FT ACT\_SITE 202 202 BINDING OF CO(2) ACTIVATES THE ENZYME.  
SQ SEQUENCE 475 AA: 52775 MW: 4HDC6186 CR032.

Query Match 45.5% Score 65: DB 8: Length 475:  
Best Local Similarity 38.9% Pred. NO. 9,08e-00:  
Matches 7: Conservative 6: Mismatches 5: Indels 0: Gaps 0:

CU 255 :FARELGVPIVNHDIITG 272  
:|||||:|:|:  
CV 1 :FSKNLNKLNMPYIAG 18

Search completed: Mon Jun 19 16:13:30 2000  
Job time : 11 secs.



10-MAR-1997: 340740.  
 10-MAR-1996: 35080702.  
 PA (MEIP) MELJI MILK PROD CO LTD.  
 PA Daitiki K, Iwara A, Kito K, Kume A, Sone T;  
 DR WPI: 97-470495/43.  
 PT Peptide immunotherapeutic agent to treat allergic diseases -  
 PT contains multi-epitope peptide containing T cell epitope regions  
 PT from different allergens  
 PT Claim 9: Page 42: 58pp; Japanese.  
 CC The present sequence represents a multi-epitope peptide which is used as  
 CC a new immunotherapeutic agent. It comprises T cell epitope regions from 2  
 CC or more different allergens (preferably linked via arginine or lysine  
 CC dimers), where the T cell epitope regions have a positivity index  
 CC greater than 1.0 as measured in a patient group responding to the  
 CC allergen; have at least 70% reactivity with lymphocytes from patients  
 CC responding to the allergen; and are not reactive with immunoglobulin E  
 CC (IgE) antibodies from patients responsive to the allergen. The agent can  
 CC be used to prevent and treat a wide variety of allergic diseases, e.g. by  
 CC desensitisation. Side effects, e.g. those mediated by IgE, are reduced.  
 CC Sequence 31 AA;

Query Match 100.0%; Score 125; DB 1; Length 31;  
 Best Local Similarity 100.0%; Pred. No. 1.10e-06;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 1 SSGKNEGNTIYNNNEAFKVE 20  
 QY 1 SSGKNEGNTIYNNNEAFKVE 20  
 IIIIIIIIIIIIIIIIIII

RESULT 3  
 ID W42121 standard; protein: 354 AA.  
 AC W42121

DE Japanese cypress pollen antigen Chaol.  
 DE Japanese cypress pollen; antigen; T-cell epitope; Chaol; Chaol2;  
 KW diagnosis; allergy; spring tree pollen disease; pollinosis.  
 KW Chamaecyparis obtusa.  
 PX W09747648-A1.  
 PX 18-DEC-1997

DE 12-JUN-1997: J:2031.  
 DE 14-JUN-1996: JP-153527.  
 PA (MEIP) MELJI MILK PROD CO LTD  
 PA Daitiki K, Kito K;  
 DR WPI: 96-052242/05.  
 PT T-cell epitope peptide portion of Japanese cypress pollen antigens  
 PT Chaol and Chaol2 used for diagnosis and treatment of spring tree  
 PT pollen disease  
 PT Example 1: Page 14-15: 7pp; Japanese.  
 CC The present sequence represents Japanese cypress pollen antigen Chaol.  
 CC The present invention describes peptides which correspond to the T-cell  
 CC epitope sites on Japanese cypress pollen antigens Chaol and Chaol2. The  
 CC peptides can be used as a reagent for the diagnosis of allergy to  
 CC Japanese cypress pollen, and as an antigen in the treatment and  
 CC prevention of spring tree pollen disease in which the pollinosis  
 CC involves reactivity to Japanese cypress pollen.  
 CC Sequence 354 AA;

Query Match 100.0%; Score 125; DB 1; Length 354;  
 Best Local Similarity 100.0%; Pred. No. 1.10e-06;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 311 SSGKNEGNTIYNNNEAFKVE 310  
 QY 1 SSGKNEGNTIYNNNEAFKVE 20  
 IIIIIIIIIIIIIIIIIII

RESULT 4  
 ID W54344 standard; Protein: 354 AA.  
 AC W04344

DE 28-NOV-1996 (first entry)  
 DE Chamaecyparis obtusa pollen allergen Cha o I (A).  
 KW Pollen allergen; Cha o I; T-cell epitope; prevention; treatment;

KW pollinosis.  
 QS Chamaecyparis obtusa.  
 PN 508176192-A.  
 PD 09-JUL-1996.  
 PF 21-DEC-1994: 335089.  
 PA (MEIP) MELJI MILK PROD CO LTD  
 DR WPI: 96-368225/37.  
 DR N-PSDB: I38519.  
 PT DNA encoding Chamaecyparis obtusa pollen allergen - T cell  
 PT epitopes) of which are useful in development of preventative and  
 PT treating agent for C. obtusa pollen pollinosis  
 PS Claim 2: Pages 10-11: 17pp; Japanese.  
 CC The present sequence is the C. obtusa pollen allergen Cha o I,  
 CC the T-cell epitopes of which can be used in the development of a  
 CC preventive and treating agent for C. obtusa pollen pollinosis.  
 CC C. obtusa pollen (2.4 kg) was decreased with diethyl ether, and  
 CC dried at room temp. overnight. Cha o I was sepd. from it and  
 CC purified. RNA was extracted from C. obtusa pollen, and mRNA and  
 CC cDNA derived.  
 CC Sequence 354 AA;

Query Match 100.0%; Score 125; DB 1; Length 354;  
 Best Local Similarity 100.0%; Pred. No. 1.10e-06;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 311 SSGKNEGNTIYNNNEAFKVE 310  
 QY 1 SSGKNEGNTIYNNNEAFKVE 20  
 IIIIIIIIIIIIIIIIIII

RESULT 5  
 ID W04345 standard; Protein: 375 AA.  
 AC W04345

DE 28-NOV-1996 (first entry)  
 DE Chamaecyparis obtusa pollen allergen Cha o I cDNA (B).  
 KW Pollen allergen; Cha o I; T-cell epitope; prevention; treatment;  
 KW pollinosis.  
 QS Chamaecyparis obtusa.  
 PN 508176192-A.  
 PD 09-JUL-1996.

PF 21-DEC-1994: 335089.  
 PA (MEIP) MELJI MILK PROD CO LTD.  
 DR WPI: 96-368225/37.  
 DR N-PSDB: I38519.  
 PT DNA encoding Chamaecyparis obtusa pollen allergen - T cell  
 PT epitopes) of which are useful in development of preventative and  
 PT treating agent for C. obtusa pollen pollinosis  
 PS Claim 2: Pages 10-11: 17pp; Japanese

CC The present sequence is the C. obtusa pollen allergen Cha o I,  
 CC the T-cell epitopes of which can be used in the development of a  
 CC preventive and treating agent for C. obtusa pollen pollinosis.  
 CC C. obtusa pollen (2.4 kg) was decreased with diethyl ether, and  
 CC dried at room temp. overnight. Cha o I was sepd. from it and  
 CC purified. RNA was extracted from C. obtusa pollen, and mRNA and  
 CC cDNA derived.  
 CC Sequence 375 AA;

Query Match 100.0%; Score 125; DB 1; Length 375;  
 Best Local Similarity 100.0%; Pred. No. 1.10e-06;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 332 SSGKNEGNTIYNNNEAFKVE 351  
 QY 1 SSGKNEGNTIYNNNEAFKVE 20  
 IIIIIIIIIIIIIIIIIII

RESULT 6  
 ID R45577 standard; Protein: 367 AA.  
 AC R45577

DE 13-JUL-1994 (first entry)  
 DE Jun 5







PR 10-JUL-1992; WO-005651.  
 PR 01-SEP-1992; US-93893C.  
 PA (IMM-) IMMUNOLOGIC PHARM CORP.  
 PI Rond JF, Garman RD, Griffith LJ, Kuo M, Pollock J.  
 DR WPI: 94-035056/24.  
 PI Antigens derived from Japanese cedar pollen allergen Cryj I -  
 PI contain at least two T cell epitope(s), used to treat or diagnose  
 PI allergy.  
 PS Claim 76; Fig 1B; 13pp; English.  
 CC The sequence is that of an isolated peptide of the Japanese cedar  
 CC pollen allergen Cryj I. The peptide, CJI-44, can be used for  
 CC the treatment and diagnosis of allergies associated with Japanese  
 CC cedar pollen. It has enhanced therapeutic properties but reduced  
 CC side effects compared to naturally occurring allergens.  
 SC Sequence 30 AA:

Query Match 74.4%; Score 93; DB 1; Length 30;  
 Best Local Similarity 70.0%; Pred. No. 6.61e-03;  
 Matches 14; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

DB 11 SSGKYEGGNIYIKKEAFNVE 30  
 QY 1 SSGKNEGTNIYNNNEAFKVE 20

## RESULT 14

ID R75244 standard; Protein: 353 AA.  
 AC R75244.  
 DE Japanese cedar pollen allergen Cryj I.  
 KW Japanese cedar pollen allergen; Cryj I; T-cell epitope; peptides;  
 KW prevention; treatment; cryptomeria pollinosis.  
 OS Cryptomeria japonica.  
 PE Key Location/Qualifiers  
 FI peptide 61..75  
 FI peptide 91..105  
 FI peptide 106..120  
 FI peptide 146..160  
 FI peptide 211..225  
 FI peptide 326..340  
 FI peptide 345..349  
 FI peptide 307119295-A.  
 PN 09-MAY-1995.  
 PR 20-OCT-1993; JP-262626.  
 PR 20-OCT-1993; JP-252626.  
 PA (MEIP) MEIJI MILK PROD CO LTD.  
 DR WPI: 95-203834/27.  
 PI New cryptomeria pollen allergen T-cell epitope peptide - used for  
 PI prevention, treatment and investigation of Japanese cedar pollinosis  
 PS Disclosure; Figs 1-2; 8pp; Japanese.  
 CC R75388 is the Japanese cedar pollen allergen Cryj I, from which the  
 CC T-cell epitope peptides R89289-R89295 were derived. The peptides  
 CC can be used for the prevention and treatment of cryptomeria  
 CC pollinosis, and also for the investigation of pollinosis.  
 SC Sequence 353 AA:

Query Match 74.4%; Score 93; DB 1; Length 353;  
 Best Local Similarity 70.0%; Pred. No. 6.61e-03;  
 Matches 14; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

DB 311 SSGKYEGGNIYIKKEAFNVE 330  
 QY 1 SSGKNEGTNIYNNNEAFKVE 20

## RESULT 15

ID R81587 standard; Protein: 353 AA.  
 AC R81587.  
 DE Cedar pollen allergen B.  
 KW Cedar pollen; allergen; immunoglobulin E; IgE; T-cell epitope;  
 KW antibody; pollinosis; therapy; immunotherapy.  
 OS Cryptomeria japonica.  
 PN EP-700929-A2.  
 PC 13-MAR-1996.  
 PR 08-SEP-1995; 306295.  
 PR 10-SEP-1994; JP-242137.  
 PR 14-JUL-1995; JP-200221.  
 PR 14-JUL-1995; JP-200204.  
 PA (HATB) HAYASHIBARA SEIBUTSU KAGAKU.  
 PI Hino K, Saito S, Taniguchi T.  
 DR WPI: 95-140976/15.  
 PI New peptide(s) derived from cedar pollen allergens - activate  
 PI allergen-specific T-cells, but not allergen-specific IgE antibodies.  
 PI used for treating cedar pollinosis.  
 PS Claim 5; Page 31-32; 36pp; English.  
 CC Synthetic peptides based on portions of cedar pollen allergens A  
 CC (R81586) and B (R81587) were tested for their ability to activate  
 CC cedar allergen-specific T-cells but not allergen-specific IgE  
 CC antibodies. 6 Peptides (R81588-R81595) were identified as T-cell  
 CC epitopes. These peptides, plus subsequences (R81573-79) essential  
 CC for T-cell recognition, and homologous peptides (R81588-95) can  
 CC be used as immunotherapeutic agents to treat or prevent cedar  
 CC pollinosis, avoiding side-effects such as anaphylaxis.  
 SC Sequence 353 AA:

Query Match 74.4%; Score 93; DB 1; Length 353;

Best Local Similarity 70.0%; Pred. No. 6.61e-03;  
Matches 14; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

DB 311 SSGKYEGGNIYIKKEAFNVE 330  
 QY 1 SSGKNEGTNIYNNNEAFKVE 20

Search completed: Mon Jun 19 16:15:59 2000  
 Job time : 6 secs.



ASSEMBLY AND USE THEREOF  
 NUMBER OF SEQUENCES: 4  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/488,316  
 FILING DATE: 05-MAR-1990  
 SEQ ID NO:4:  
 LENGTH: 1365  
 NO SEQUENCE 1365 AA: 156437 MW: 9878636 CN:

Query Match 44.0% Score 55: DB 5: Length 1365:  
 Best Local Similarity 46.7%: Pred. No. 3,394-01:  
 Matches 7: Conservative 3: Mismatches 5: Indels 0: Gaps 0:

Db 840 EGTNIYNNNEAFKVE 20  
 QY 6 EGTNIYNNNEAFKVE 20

RESULT 3  
 ID US-08-616-241-1 STANDARD: PRT: 190 AA:

XX xxxxxx

Sequence 1. Application US/08816241

Sequence 1. Application US/08816241

Patent No. 5934,185

GENERAL INFORMATION:

APPLICANT: Bandman, Gilda

APPLICANT: Goli, Surya K.

TITLE OF INVENTION: NOVEL RNA EDITING ENZYME

NUMBER OF SEQUENCES: 5

CORRESPONDENCE ADDRESS:

ADDRESSEE: Inocyte Pharmaceuticals, Inc.

STREET: 3174 Porter Drive

CITY: Palo Alto

STATE: CA

COUNTRY: USA

ZIP: 94304

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FASTA for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/816,241

FILING DATE: Filed Herewith

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Billings, Lucy J.

REGISTRATION NUMBER: 36,749

REFERENCE/DOCKET NUMBER: PF-0239 US

TELECOMMUNICATION INFORMATION:

TELEPHONE: 415-855-0555

TELEFAX: 415-845-4166

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 190 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

IMMEDIATE SOURCE:

LIBRARY: PROSTUI09

CLONE: 1645823

NO SEQUENCE 190 AA: 22826 MW: 191242 CN:

Query Match 44.0% Score 55: DB 1: Length 190:  
 Best Local Similarity 66.7%: Pred. No. 4,31e-01:

Matches 6: Conservative 3: Mismatches 0: Indels 0: Gaps 0:  
 Db 161 VYNDNEPK 169  
 QY 10 IYNNNEAFK 18

RESULT 4  
 ID US-09-568-459A-12 STANDARD: PRT: 2710 AA:

XX xxxxxx

Sequence 12. Application US/08568459A

Sequence 12. Application US/08568459A

Patent No. 5849305

GENERAL INFORMATION:

APPLICANT: Sim, Kim I.

APPLICANT: Chitnis, Chetan

APPLICANT: Miller, Louis R.

APPLICANT: Peterson, David S.

APPLICANT: Sr, Xin-zhaun

APPLICANT: Wellens, Thomas E.

TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX

TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS

NUMBER OF SEQUENCES: 37

CORRESPONDENCE ADDRESS:

ADDRESSEE: Knobbe Martens Olson & Bear

STREET: 620 Newport Center Drive 19th Floor

CITY: Newport Beach

STATE: California

COUNTRY: US

ZIP: 92650

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #10, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/568,459A

FILING DATE: 07-DEC-1995

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Knoppe Olson & Bear

REGISTRATION NUMBER: 36,749

REFERENCE/DOCKET NUMBER: N011110001

TELECOMMUNICATION INFORMATION:

TELEPHONE: (619) 235-4550

TELEFAX: (619) 235-0176

INFORMATION FOR SEQ ID NO: 12:

SEQUENCE CHARACTERISTICS:

LENGTH: 2710 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

HYPOTHETICAL: NO

ORIGINAL SOURCE:

ORGANISM: Plasmodium falciparum

NO SEQUENCE 2710 AA: 308038 MW: 37145344 CN:

Query Match 44.0% Score 55: DB 2: Length 2710:  
 Best Local Similarity 38.5%: Pred. No. 4,31e+01:

Matches 5: Conservative 5: Mismatches 3: Indels 0: Gaps 0:

Db 2205 EGTKIFDDDKTEK 2217

QY 6 EGTNIYNNNEAFK 18

RESULT 5



STATE: Colorado  
COUNTRY: U.S.A.  
ZIP: 80203  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/580.726A  
FILING DATE: 12-JUL-1996  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Concell, Gary C.  
REGISTRATION NUMBER: 32,020  
REFERENCE/DOCKET NUMBER: 2618-46-C1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (303) 863-9700  
TELEFAX: (303) 563-0223  
INFORMATION FOR SEQ ID NO: 58:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 522 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: Protein  
SEQUENCE 522 AA: 59930 MW: 151653 CN:  
Query Match 42.4% Score 53 DB 1 Length 522:  
Best Local Similarity 35.3% Pred. No. 6,950-01:  
Matches 6: Conservative 7: Mismatches 4: Indels 0: Gaps 0:  
DB 161 KNYSHVFPVNDSEKIE 207  
1 KNYSHVFPVNDSEKIE 207  
4 KNEGTNIYNNNEAFK 20  
RESULT 8  
DE PCT-US93-11703-75 STANDARD: PRI: 1018 AA:  
XX XXXXX  
Sequence 75, Application PCT/US93/11703  
Sequence 75, Application PCT/US93/11703  
GENERAL INFORMATION:  
APPLICANT: Chiron Mimotopes Pty Ltd.  
TITLE OF INVENTION: Local Effector  
NUMBER OF SEQUENCES: 75  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Grant D. Green,  
STREET: 4560 Horton St.,  
CITY: Emeryville  
STATE: CA  
COUNTRY: USA  
ZIP: 94608  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US93/11703  
FILING DATE: 28-DEC-1993  
CLASSIFICATION:  
PRIOR APPLICATION NUMBER:  
APPLICATION NUMBER: US 07/984,852  
FILING DATE: 02-DEC-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Green, Grant D.  
REGISTRATION NUMBER: 31,259  
REFERENCE/DOCKET NUMBER: 0222.10:

TELECOMMUNICATION INFORMATION:  
TELEPHONE: 510-601-2706  
TELEFAX: 510-655-3542  
INFORMATION FOR SEQ ID NO: 75:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1018 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
SEQUENCE 1018 AA: 11493 MW: 5251154 CN:  
Query Match 42.4% Score 53 DB 4 Length 1018:  
Best Local Similarity 40.0% Pred. No. 6,950-01:  
Matches 6: Conservative 5: Mismatches 4: Indels 0: Gaps 0:  
DB 453 RGGTINVNDTEVER 467  
QY 4 KNEGTNIYNNNEAFK 18  
RESULT 9  
DE US-08-483-857-10 STANDARD: PRI: 261 AA:  
XX XXXXX  
Sequence 10, Application US/08483857  
Sequence 10, Application US/08483857  
Patent No. 6020125  
GENERAL INFORMATION:  
APPLICANT: Grant, Voon Loong  
TITLE OF INVENTION: BASAL BODY PROTEIN GENES OF  
TITLE OF INVENTION: CANCER TREATMENT  
NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sim & McBratney  
STREET: Suite 701, 330 University Avenue  
CITY: Toronto  
STATE: Ontario  
COUNTRY: Canada  
ZIP: M5S 1R7  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/8/484,857  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Stewart, Michael L.  
REGISTRATION NUMBER: 24,973  
REFERENCE/DOCKET NUMBER: 1018-404  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (416) 595-1155  
TELEFAX: (416) 595-1163  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 261 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
SEQUENCE 261 AA: 27650 MW: 357696 CN:  
Query Match 40.8% Score 51 DB 3 Length 261:  
Best Local Similarity 33.3% Pred. No. 1,116+02:  
Matches 5: Conservative 5: Mismatches 5: Indels 0: Gaps 0:



Sequence 21, Application US/08467974  
Patent No. 5965385  
GENERAL INFORMATION:  
APPLICANT: READ, Randy J.  
APPLICANT: STEIN, Penelope E.  
APPLICANT: COCKLE, Stephen A.  
APPLICANT: COMEN, Raymond P.  
APPLICANT: LOOSMORE, Sheena  
APPLICANT: KLEIN, Michel H.  
APPLICANT: ARMSTRONG, Glen D.  
APPLICANT: HAZES, Bart  
TITLE OF INVENTION: MODIFICATION OF PERTUSSIS TOXIN  
NUMBER OF SEQUENCES: 46  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sim & McBurney  
STREET: Suite 701, 330 University Avenue  
CITY: Toronto  
STATE: Ontario  
COUNTRY: Canada  
ZIP: M5G 1R7  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/467,974  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/467,536  
FILING DATE: 06-JUN-1995  
APPLICATION NUMBER: US 08/292,958  
FILING DATE: 22-AUG-1994  
APPLICATION NUMBER: US 08/251,121  
FILING DATE: 31-MAY-1994  
APPLICATION NUMBER: US 08/110,947  
FILING DATE: 24-AUG-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: STEWART, Michael I.  
REGISTRATION NUMBER: 24,373  
REFERENCE/DOCKET NUMBER: 1038-454 MIS.vg  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (416) 595-1153  
TELEFAX: (416) 595-1153  
INFORMATION FOR SEQ ID NO: 21:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 69 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
SEQUENCE 69 AA: 7591 MW: 25533 CN:

Query Match 40.0% Score 50: DB 2: Length 69:  
Best Local Similarity 44.4% Pred. No. 1.41e+02:  
Matches 8: Conservative 6: Mismatches 3: Indels 1: Gaps 1:  
Db 6 TGRVEYTK-YNDDDTFTV 22  
||| | | | | | | | |  
2 SGKNEGTNIYNNNEAFKV 19

RESULT 13  
ID US-08-467-536-21 STANDARD: PRT: 69 AA.  
XX  
AC xxxxxx  
XX  
DT  
XX  
DE Sequence 21, Application US/08467536  
XX Sequence 21, Application US/08467536  
XX Patent No. 6018022

Sequence 21, Application US/08467536  
Patent No. 5977304  
GENERAL INFORMATION:  
APPLICANT: READ, Randy J.  
APPLICANT: STEIN, Penelope E.  
APPLICANT: COCKLE, Stephen A.  
APPLICANT: COMEN, Raymond P.  
APPLICANT: LOOSMORE, Sheena  
APPLICANT: KLEIN, Michel H.  
APPLICANT: ARMSTRONG, Glen D.  
APPLICANT: HAZES, Bart  
TITLE OF INVENTION: MODIFICATION OF PERTUSSIS TOXIN  
NUMBER OF SEQUENCES: 46  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sim & McBurney  
STREET: Suite 701, 330 University Avenue  
CITY: Toronto  
STATE: Ontario  
COUNTRY: Canada  
ZIP: M5G 1R7  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/467,536  
FILING DATE: 06-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/292,958  
FILING DATE: 22-AUG-1994  
APPLICATION NUMBER: US 08/251,121  
FILING DATE: 31-MAY-1994  
APPLICATION NUMBER: US 08/110,947  
FILING DATE: 24-AUG-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: STEWART, Michael I.  
REGISTRATION NUMBER: 24,373  
REFERENCE/DOCKET NUMBER: 1038-454 MIS.vg  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (416) 595-1153  
TELEFAX: (416) 595-1153  
INFORMATION FOR SEQ ID NO: 21:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 69 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
SEQUENCE 69 AA: 7591 MW: 25533 CN:

Query Match 40.0% Score 50: DB 2: Length 69:  
Best Local Similarity 44.4% Pred. No. 1.41e+02:  
Matches 8: Conservative 6: Mismatches 3: Indels 1: Gaps 1:  
Db 6 TGRVEYTK-YNDDDTFTV 22  
||| | | | | | | | |  
2 SGKNEGTNIYNNNEAFKV 19

RESULT 14  
ID US-08-467-976-21 STANDARD: PRT: 69 AA.  
XX  
AC xxxxxx  
XX  
DT  
XX  
DE Sequence 21, Application US/08467976  
XX Sequence 21, Application US/08467976  
XX Patent No. 6018022

CC GENERAL INFORMATION:  
CC APPLICANT: READ, Randy J.  
CC APPLICANT: STEIN, Petelope E.  
CC APPLICANT: COCKLE, Stephen A.  
CC APPLICANT: COHEN, Raymond P.  
CC APPLICANT: LOOSMORE, Sheena  
CC APPLICANT: KLEIN, Michael H.  
CC APPLICANT: ARMSTRONG, Glen D.  
CC APPLICANT: HAZES, Bart  
CC TITLE OF INVENTION: MODIFICATION OF PERTUSSIS TOXIN  
CC NUMBER OF SEQUENCES: 46  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Sir & McBurney  
CC STREET: Suite 701, 330 University Avenue  
CC City: Toronto  
CC STATE: Ontario  
CC COUNTRY: Canada  
CC ZIP: M5G 1R7  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: Patent In Release #1.0, Version #1.25  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/457,976  
CC FILING DATE: 06-JUN-1995  
CC CLASSIFICATION: 435  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 08/292,968  
CC FILING DATE: 22-AUG-1994  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 08/251,121  
CC FILING DATE: 31-MAY-1994  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 08/110,947  
CC FILING DATE: 24-AUG-1993  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: STEWART, Michael I.  
CC REGISTRATION NUMBER: 24,973  
CC REFERENCE/DOCKET NUMBER: 1038-453 M 51vg  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (416) 595-1155  
CC TELEFAX: (416) 595-1153  
CC INFORMATION FOR SEQ ID NO: 21:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 69 amino acids  
CC TYPE: amino acid  
CC STRANDEDNESS: single  
CC TOPOLOGY: linear  
CC SEQUENCE 69 AA: 7691 MW: 25533 CN:  
  
Db 6 TGKVEYK.YNDDTFTV 22  
QY 2 SKNEGTYNNNEAFKV 19  
  
RESULT 15  
IL US-08-937-972-3 STANDARD: PRT: 321 AA.  
XX  
AC xxxxxx  
XX  
DE  
XX  
DE  
XX  
XX  
Sequence 3, Application US/08937972  
Sequence 3, Application US/08937972  
Patent No. 5932443  
CC GENERAL INFORMATION:  
CC APPLICANT: Lal, Preeti

CC APPLICANT: Bandman, Olga  
CC APPLICANT: Corley, Neil C.  
CC APPLICANT: Shah, Purvi  
CC TITLE OF INVENTION: ANTIGENS  
CC NUMBER OF SEQUENCES: 6  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Incyte Pharmaceuticals, Inc.  
CC STREET: 3174 Porter Drive  
CC City: Palo Alto  
CC STATE: CA  
CC COUNTRY: USA  
CC ZIP: 94304  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Diskette  
CC COMPUTER: IBM Compatib.e  
CC OPERATING SYSTEM: DOS  
CC SOFTWARE: FastSeq for Windows Version 2.0  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/937,972  
CC FILING DATE: Herewith  
CC CLASSIFICATION: 424  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER:  
CC FILING DATE:  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Billings, Lucy J.  
CC REGISTRATION NUMBER: 35,749  
CC REFERENCE/DOCKET NUMBER: PF-0400 US  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: 650-855-0555  
CC TELEFAX: 650-845-4156  
CC TELEX:  
CC INFORMATION FOR SEQ ID NO: 3:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 321 amino acids  
CC TYPE: amino acid  
CC STRANDEDNESS: single  
CC TOPOLOGY: linear  
CC IMMEDIATE SOURCE:  
CC LIBRARY: BLADNOT04  
CC CLONE: 13:8190  
CC SEQUENCE 321 AA: 35892 MW: 544482 FN:  
  
Query Match 40.0% Score 691 IP: 40 Length 321  
Best Local Similarity 30.8% Pred No. 14002  
Matches 4: Conservative 60 Mismatches 3: Indels 0: Caps 0:  
  
Db 171 INVGPHENFNE 183  
QY 8 TNYNNNEAFKE 20  
  
Search completed: Tue Jun 20 13:37:07 2000  
Job time : 6 secs.













```

XX Sequence 57, Application US/08350225
CC GENERAL INFORMATION:
CC APPLICANT: Griffith, Irwin J.
CC APPLICANT: Pollock, Joanne.
CC APPLICANT: Bond, Julian F.
CC APPLICANT: Garman, Richard D.
CC APPLICANT: Kuo, Mei-Chang.
CC APPLICANT: Yeung, Siu-mei H.
CC APPLICANT: Brauer, Andrew.
CC APPLICANT: Exley, Mark A.
CC APPLICANT: Powers, Steven P.
CC TITLE OF INVENTION: Allergic Proteins And Peptides From
CC TITLE OF INVENTION: Japanese Cedar Pollen
CC NUMBER OF SEQUENCES: 261
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Immunologic Pharmaceutical Corporation, Inc.
CC STREET: 610 Lincoln St
CC CITY: Waltham
CC STATE: MA
CC COUNTRY: USA
CC ZIP: 02154
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/458,940
CC FILING DATE: December 6, 1994
CC CLASSIFICATION: 424
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/426,249
CC FILING DATE: April 8, 1994
CC APPLICATION NUMBER: 07/938,990
CC FILING DATE: September 1, 1992
CC APPLICATION NUMBER: PCT/US93/00139
CC FILING DATE: January 15, 1993
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Darlene A. Vanstone
CC REGISTRATION NUMBER: 35,729
CC REFERENCE/DOCKET NUMBER: 025.6 US (IMI-028CP2)
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (617) 466-6800
CC TELEFAX: (617) 466-6840
CC INFORMATION FOR SEQ ID NO: 57:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 20 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC FRAGMENT TYPE: internal
CC SEQ SEQUENCE 20 AA: 2221 MW: 2174 CN:

Query Match 74.4% Score 93; DB 10; Length 20;
Best Local Similarity 70.0%; Pred. No. 4.09e-03;
Matches 14; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

DB 1 SSGKYEGGNTYTKKEAFNVE 20
QY 1 SSGKNEGTNIYNNNEAFKVE 20

RESULT 12
ID US-08-350-225-57 STANDARD: PRT: 20 AA.
XX AC xxxxxx
XX DT
XX DE
XX Sequence 57, Application US/08350225
CC Sequence 57, Application US/08226248A
CC Sequence 57, Application US/08226248A
CC Sequence 57, Application US/08350225

```

```

CC GENERAL INFORMATION:
CC APPLICANT: Griffith, Irwin J.;
CC APPLICANT: Pollock, Joanne;
CC APPLICANT: Bond, Julian F.;
CC APPLICANT: Garman, Richard D.;
CC APPLICANT: Kuo, Mei-Chang;
CC APPLICANT: Yeung, Siu-mei H.;
CC APPLICANT: Brauer, Andrew;
CC APPLICANT: Exley, Mark A.;
CC APPLICANT: Powers, Steven P.;
CC TITLE OF INVENTION: Allergenic Proteins And Peptides From
CC TITLE OF INVENTION: Japanese Cedar Pollen
CC NUMBER OF SEQUENCES: 201
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Immunologic Pharmaceutical Corporation, Inc.
CC STREET: 610 Lincoln St
CC CITY: Waltham
CC STATE: MA
CC COUNTRY: USA
CC ZIP: 02154
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/09/226,248A
CC FILING DATE: April 8, 1994
CC CLASSIFICATION: 424
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 07/938,990
CC FILING DATE: September 1, 1992
CC APPLICATION NUMBER: PCT/US93/00139
CC FILING DATE: January 15, 1993
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Darlene A. Vanstone
CC REGISTRATION NUMBER: 35,729
CC REFERENCE/DOCKET NUMBER: 025.5 US (IM1-028CP)
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (617) 466-6000
CC TELEFAX: (617) 466-6040
CC INFORMATION FOR SEQ ID NO: 57:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 20 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC FRAGMENT TYPE: internal
CC SEQUENCE: 20 AA: 2221 MW: 2174 CN:

Db Query Match 74.4% Score 93: DB 7: Length 20;
Best Local Similarity 70.0% Pred. No. 4: 09e-03;
Matches 14: Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 SSGKVEGGNIYTKKEAFNVE 20
|||||.....|
RESULT 14
ID US-08-226-248A-91 STANDARD: PRT: 26 AA.
XX
AC xxxxxx
DE
DT
XX
XX Sequence 91, Application US/08226248A
CC Sequence 91, Application US/08226248A
CC GENERAL INFORMATION:
CC APPLICANT: Griffith, Irwin J.;
CC APPLICANT: Pollock, Joanne;
CC APPLICANT: Bond, Julian F.;

```

```

CC APPLICANT: Garman, Richard D.;
CC APPLICANT: Kuo, Mei-Chang;
CC APPLICANT: Yeung, Siu-mei H.;
CC APPLICANT: Brauer, Andrew;
CC APPLICANT: Exley, Mark A.;
CC APPLICANT: Powers, Steven P.;
CC TITLE OF INVENTION: Allergenic Proteins And Peptides From
CC TITLE OF INVENTION: Japanese Cedar Pollen
CC NUMBER OF SEQUENCES: 201
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Immunologic Pharmaceutical Corporation, Inc.
CC STREET: 610 Lincoln St
CC CITY: Waltham
CC STATE: MA
CC COUNTRY: USA
CC ZIP: 02154
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/09/226,248A
CC FILING DATE: April 8, 1994
CC CLASSIFICATION: 424
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 07/938,990
CC FILING DATE: September 1, 1992
CC APPLICATION NUMBER: PCT/US93/00139
CC FILING DATE: January 15, 1993
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Darlene A. Vanstone
CC REGISTRATION NUMBER: 35,729
CC REFERENCE/DOCKET NUMBER: 025.5 US (IM1-028CP)
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (617) 466-6000
CC TELEFAX: (617) 466-6040
CC INFORMATION FOR SEQ ID NO: 91:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 26 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC FRAGMENT TYPE: internal
CC SEQUENCE: 26 AA: 2873 MW: 3771 CN:

Db Query Match 74.4% Score 93: DB 7: Length 26;
Best Local Similarity 70.0% Pred. No. 4: 09e-03;
Matches 14: Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 SSGKNEGTNIYNNNEAFKVE 20
|||||.....|
RESULT 15
ID US-08-467-697-91 STANDARD: PRT: 26 AA.
XX
AC xxxxxx
DE
DT
XX
XX Sequence 91, Application US/08467697
CC Sequence 91, Application US/08467697
CC GENERAL INFORMATION:
CC APPLICANT: Griffith, Irwin J.;
CC APPLICANT: Pollock, Joanne;
CC APPLICANT: Bond, Julian F.;
CC APPLICANT: Garman, Richard D.;
CC APPLICANT: Kuo, Mei-Chang;
CC APPLICANT: Yeung, Siu-mei H.;
CC APPLICANT: Brauer, Andrew;

```

APPLICANT: Exley, Mark A.;  
 APPLICANT: Powers, Steven P.  
 TITLE OF INVENTION: Allergenic Proteins And Peptides From  
 TITLE OF INVENTION: Japanese Cedar Pollen  
 NUMBER OF SEQUENCES: 261  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Immunologic Pharmaceutical Corporation, Inc.  
 STREET: 610 Lincoln St  
 CITY: Waltham  
 STATE: MA  
 COUNTRY: USA  
 ZIP: 02154  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patulin Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/467,697  
 FILING DATE: June 6, 1995  
 CLASSIFICATION: 424  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 08/350,225  
 FILING DATE: December 6, 1994  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Jane E. Remillard  
 REGISTRATION NUMBER: 38,872  
 REFERENCE/DOCKET NUMBER: 025.6 USD4 (TM-028GPD4)  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (617) 227-7400  
 TELEFAX: (617) 227-5941  
 INFORMATION FOR SEQ ID NO: 91:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 26 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: Peptide  
 FRAGMENT TYPE: Internal  
 SEQ SEQUENCE 26 AA: 2873 MW: 3776 CN:

Query Match 74.4% Score 93; DB 10; Length 26;  
 Best Local Similarity 70.0%; Pref. No. 4.09e-03;  
 Matches 14; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

06 7 SSGKYECUN:YTKKAPNVE 26  
 07 1 SSGKNESTNYNNNEAPKVE 26

Search completed: Mon Jun 19 16:26:31 2000  
 Job time : 17 secs.







```

ALTERNATE_NAMES  protein U9470.8; protein YLR182W
ORGANISM          Saccharomyces cerevisiae
DATE              31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change
                  18-Jun-1999
ACCESSIONS       S3316; S51427
REFERENCE        S3316;
                  Breeders, J.; Nasmyth, K.
                  Nature (1987) 329:651-654
                  Similarity between cell-cycle genes of budding yeast and
                  fission yeast and the Notch gene of Drosophila.
                  #cross-references MIM:88014241
                  #accession S3316;
                  #molecule_type DNA
                  #residues 1-803 #label BRE
                  #cross-references EMBL:X06238; NID:g4597; PIDN:CAA2581.1; PID:g4598
REFERENCE        S51414
                  Wohldmann, P.
                  #submission submitted to the EMBL Data Library, November 1994
                  #description The sequence of S. cerevisiae cosmid 9470.
                  #accession S51427
                  #molecule_type DNA
                  #residues 1-803 #label WOH
                  #cross-references EMBL:U17246; NID:g571192; PIDN:AAB67450.1;
                  #cross-references MIM:9577200; MIPS:YLR182W
GENETICS
#gene            SGD:SWI6; SDS11; PSL8
#map_position 12R
#cross-references SGD:S0004172; MIPS:YLR182W
CLASSIFICATION  #superfamily regulatory protein SWI6; ankyrin repeat homology
KEYWORDS         DNA binding; transcription regulation
FEATURE          317-349
                  #domain ankyrin repeat homology #label AN1\
                  #domain ankyrin repeat homology #label AK2
                  469-501
                  #length 803 #molecular-weight 90559 #checksum 7241
SUMMARY
Query Match      48.08; Score 60; DB 1; Length 803;
Best Local Similarity 45.08; Pred. No. 3,38e+00;
Matches          9; Conservative 6; Mismatches 5; Indels 0; Gaps 0;
DB              651 SSGIDWSSVDADPEKVE 670
CY              1 SSGKNEGNIYNNNEAFKE 20

RESULT          6
ENTRY           671259 #type complete
TITLE           Probable flagellar basal-body rod protein (flgB1) - syphilis
                Spirochete
                #formal_name Treponema pallidum subsp. pallidum #common_name
                syphilis spirochete
                #formal_name Treponema pallidum subsp. pallidum #common_name
                syphilis spirochete
                24-Jul-1998 #sequence_revision 24-Jul-1998 #text_change
                C71259
                A71250
                Fraser, C.M.; Norris, S.J.; Weinstock, G.M.; White, G.;
                Sutton, G.G.; Dodson, R.; Gwin, M.; Hickey, E.K.; Clayton,
                R.; Ketchum, K.A.; Sodergren, E.; Hardham, J.M.; McLeod,
                M.P.; Salzberg, S.; Peterson, J.; Khalak, H.; Richardson,
                D.; Howell, J.K.; Chidambaram, M.; Utterback, T.; McDonald,
                L.; Artlich, P.; Bowman, C.; Cotton, M.D.; Fujii, C.;
                Garland, S.; Hatch, B.; Horst, K.; Roberts, K.; Watthey,
                L.; Weidman, J.; Smith, H.O.; Venter, J.C.
                Science (1998) 281:375-388
                Complete genome sequence of Treponema pallidum, the syphilis
                spirochete.
                #cross-references MIM:98332770
                #accession C71259
                #status preliminary; nucleic acid sequence not shown;
                translation not shown;
                #molecule_type DNA
                #residues 1-264 #label COL
                #cross-references GB:AE001264; GB:AE000520; NID:g3323278;
                PIDN:AAC659.5.1; PID:g3323281

```

```

#experimental_source strain Nichols
GENETICS
#gene            TP0960
CLASSIFICATION  #superfamily rod protein flgF
SUMMARY          #length 264 #molecular-weight 28597 #checksum 5918
Query Match      47.28; Score 59; DB 2; Length 264;
Best Local Similarity 40.04; Pred. No. 4,54e+00;
Matches          6; Conservative 4; Mismatches 5; Indels 0; Gaps 0;
DB              109 DGTAYATGSGSEKVD 123
CY              6 EGTNINNNNEAFKE 20

RESULT          7
ENTRY           T15906 #type complete
TITLE           Hypothetical: protein E0416.3 - Caenorhabditis elegans
                #formal_name Caenorhabditis elegans
                20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change
                T15906
ACCESSIONS       T15906
REFERENCE        T18427
                  Pauley, A.
                  #submission submitted to the EMBL Data Library, June 1995
                  #description The sequence of C. elegans cosmid 504F6.
                  #accession T15906
                  #status preliminary; translated from GB/EMBL/DDBJ
                  #molecule_type DNA
                  #residues 1-238 #label PAU
                  #cross-references EMBL:U23943; NID:g861333; PID:g661336;
                  PIDN:AAA68358.1; CDS:P04F6.3
                  #experimental_source strain Bristol N2
GENETICS
#gene            GESP:E04F6.3
#introns         30/1; 77/2; 236/3
#summary         #length 298 #molecular-weight 3272; #checksum 4744
Query Match      47.24; Score 59; DB 2; Length 298;
Best Local Similarity 36.84; Pred. No. 4,64e+00;
Matches          7; Conservative 7; Mismatches 5; Indels 0; Gaps 0;
DB              32 ARAKEDLCYVENHDEPKV 50
CY              1 SSGKNEGNIYNNNEAFKV 19

RESULT          8
ENTRY           A40550 #type complete
TITLE           Peroxisomal membrane protein, PAS3 - Yeast (Saccharomyces
                cerevisiae)
                #formal_name Saccharomyces cerevisiae
                17-Jul-1992 #sequence_revision 17-Jul-1992 #text_change
                05-Dec-1997
                A40550; S59794
                Hoehfeld, J.; Veenhuis, M.; Kunau, W.H.
                J. Cell Biol. (1991) 114:1167-1178
                #journal PAS3, a Saccharomyces cerevisiae gene encoding a peroxisomal
                integrin membrane protein essential for peroxisome
                biogenesis.
                #cross-references MIM:91373453
                #accession A40550
                #molecule_type DNA
                #residues 1-44; #label HOE
                #cross-references GB:X58407; NID:g4102; PID:g4103
                S59783
                #authors Du, Z.
                #submission submitted to the EMBL Data Library, July 1995
                #description The sequence of S. cerevisiae cosmid 9798.
                #accession S59794
                #molecule_type DNA

```

```

##residues 1-441 ##label DZ
##cross-references EMBL:U032517; NID:q914989; PID:q15004; MIPS:YKR329c
##experimental_source strain S288C (A8972)

GENETICS
#gene SGD:PEX3; PAS3
#map_position 4P
#cross-references MIPS:YKR329c; SGD:S06C2737
#feature
#feature 23-39 #domain transmembrane #status predicted #label IMN
#feature 236-152 #domain transmembrane #status predicted #label IM2
#summary
#length 441 #molecular_weight 5675 #checksum 166

Query Match 47.2% Score 59; DB 2; Length 441;
Best Local Similarity 46.7%; Pred. No. 4,64e+00;
Matches 7; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

DB 212 SKKGEVYVINEQAF 226
|||||
QY 3 KNEGTINYNNEAF 17

#result 7
#entry 264030 #type complete
#title Hypothetical protein H11458 - Haemophilus influenzae (strain
#organism RG KW20;
#formal_name Haemophilus influenzae
#date 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change
30-Jun-1998
#accessions Q64330
#reference A64000
#authors Kirschner, R.D.; Adams, M.D.; White, O.; Clayton, F.A.;
Kirkness, E.F.; Korfavade, A.R.; Bult, C.A.; Tomb, J.F.;
Dougherty, B.A.; Merrick, J.M.; McKenney, K.; Sutton, G.;
FitzHugh, W.; Fields, C.; Gocayne, J.D.; Scott, J.;
Shirley, R.; Liu, L.; Glodek, A.; Kelley, J.M.; Weidman,
J.E.; Phillips, C.A.; Spriggs, T.; Hedblom, E.; Cotton,
M.D.; Gierbach, T.; Hanna, M.G.; Nguyen, D.T.; Saudek,
D.M.; Brandon, R.C.; Fife, L.D.; Fritchman, J.B.; Furmann,
J.L.; Geohagen, N.S.M.; Gnehm, C.L.; McDonald, L.A.;
Small, K.V.; Fraser, C.M.; Smith, R.O.; Venter, J.C.
#journal Science (1995) 269:496-512
#title Whole-genome random sequencing and assembly of Haemophilus
influenzae RG
#cross-references MIM:19530636
#accession Q64030
#status nucleic acid sequence not shown; translation not shown
#molecule_type DNA
#residues 1-441 ##label DZ
#cross-references G3:U032517; G3:U142023; NID:q1574302;
PID:q1574302;
H11458

GENETICS
#start_codon GTG
#summary
#length 80 #molecular_weight 9400 #checksum 6360

Query Match 46.4% Score 58; DB 2; Length 80;
Best Local Similarity 42.1%; Pred. No. 6,97e+00;
Matches 8; Conservative 7; Mismatches 4; Indels 0; Gaps 0;

DB 28 NKNSTNYGNDSLIE 46
|||||
QY 2 SKNENGINYNNEAF 20

#result 10
#entry S36017 #type complete
#title finger protein trantrack-p69 (tkk) - fruit fly (Drosophila
melanogaster)
#organism
#formal_name Drosophila melanogaster
#date 18-Jan-1995 #sequence_revision 13-Jan-1995 #text_change
24-Sep-1998
#accessions S36017; S10881
#reference S36017

```

```

#authors Montell, C.
#submission Submitted to the EMBL Data Library, April 1993
#accession S36017
#status preliminary
#molecule_type mRNA
#residues 1-643 ##label MGN
#cross-references EMBL:X71627; NID:q297496; PID:q297497
#reference S10881
#authors Harrison, S.D.; Travers, A.A.
#journal EMBO J. (1990) 9:207-216
#title The trantrack gene encodes a Drosophila finger protein that
interacts with the 17z transcriptional regulatory region
and shows a novel embryonic expression pattern.
#cross-references MIM:19107945
#accession S10881
#status preliminary
#molecule_type DNA
#residues 3-643 ##label HAP
#cross-references EMBL:X7121; NID:q2749; PID:q8747
#note the authors translated the codon GGC for residue 73 as
Phe

GENETICS
#gene FlyBase:tkk
#cross-references FlyBase:FBgn0003870
#classification #superfamily POZ domain family
#keywords DNA binding; nucleic acid transcription regulation
#feature
#feature 19-130 #domain POZ domain homolog #label PZ
#summary
#length 643 #molecular_weight 68770 #checksum 6107

Query Match 46.4% Score 58; DB 2; Length 643;
Best Local Similarity 54.5%; Pred. No. 6,97e+00;
Matches 6; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

DB 446 DYNND3FKL 456
|||||
QY 9 NYNNNEAFV 19

#result 11
#entry T03992 #type fragment
#title Hypothetical protein T03992 - Arabidopsis thaliana
(fragment)
#organism Arabidopsis thaliana #common_name mouse ear
cress
#date 30-Apr-1994 #sequence_revision 1; #accession T03992;
T03992
#accessions T03992
#reference Z15164
#authors Bewick, M.; Murphy, G.; Allgeier, H.; Hudson, S.; Rammstedt, J.;
Mewes, H.W.; Meyer, R.F.X.; Schaefer, C.
#submission Submitted to the Protein Sequence Database, March 1993
#accession T03992
#molecule_type DNA
#residues 1-185 ##label BVV
#cross-references EMBL:U04248;
#experimental_source cultivar Col-0; #accession T03992;
T03992

GENETICS
#map_position 4
#note Intronic positions not resolved; T5L19.10
#summary
#length 185 #checksum 902

Query Match 45.6% Score 57; DB 2; Length 185;
Best Local Similarity 40.0%; Pred. No. 1,04e+01;
Matches 6; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

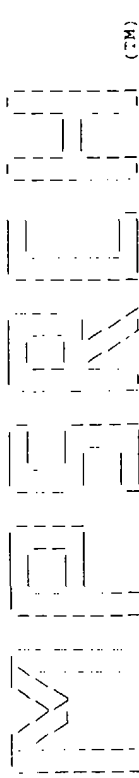
DB 102 KNEGVYNDKIFR 116
|||||
QY 4 KNEGTINYNNEAF 18

#result 12
#entry E70196 #type complete

```







Release 3.1A John F. Collins, Biocomputing Research Unit,  
Copyright (C) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

protein - protein database search, using Smith-Waterman algorithm  
Run on: Mon Jun 19 16:14:36 2000; MasPar time 3.71 Seconds  
Tabular output not generated. 164.264 Million cell updates/sec

Title: >US-09-142-524A-12  
Description: (1-20) from US09142524A.pep  
Perfect Score: 125  
Sequence: 1 SSGKNEGINYNNNEAFKVE 20

Scoring table: PAM 150  
Gap 15

Searched: 53957 seqs, 30454573 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: swiss-prot38  
1:swissprot

Statistics: Mean 27.434; Variance 35.196; scale 0.780

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed.  
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match % | Length | ID | Description | Pred. No.                      |
|------------|-------|---------------|--------|----|-------------|--------------------------------|
| 1          | 93    | 74.4          | 374    | 1  | SBP_CRYJA   | SUGI BASIC PROTEIN PRE         |
| 2          | 68    | 54.4          | 623    | 1  | TKT_HUMAN   | TRANSKETOLASE (EC 2.2.1.11)    |
| 3          | 64    | 51.2          | 623    | 1  | TKT_MOUSE   | TRANSKETOLASE (EC 2.2.1.11)    |
| 4          | 64    | 51.2          | 623    | 1  | TKT_RAT     | TRANSKETOLASE (EC 2.2.1.11)    |
| 5          | 60    | 48.0          | 803    | 1  | SW16_YEAS   | REGULATORY PROTEIN SWI         |
| 6          | 60    | 48.0          | 843    | 1  | CYAA_AERY   | ADENYLATE CYCLASE (EC 9.4.4.1) |
| 7          | 59    | 47.2          | 441    | 1  | PEX3_YEAS   | PEROXISOMAL MEMBRANE P         |
| 8          | 58    | 46.4          | 80     | 1  | YESS_HAEI   | HYPOTHETICAL PROTEIN H         |
| 9          | 58    | 46.4          | 641    | 1  | TKTB_DROME  | TRANSTRACK PROTEIN, BET        |
| 10         | 57    | 45.6          | 265    | 1  | FLGG_BORBU  | FLAGELLAR BASAL-BODY R         |
| 11         | 57    | 45.6          | 344    | 1  | ABIC_LACLA  | ABORTIVE PHAGE RESISTA         |
| 12         | 57    | 45.6          | 526    | 1  | CAP_YEAS    | ADENYLATE CYCLASE ASSOC        |
| 13         | 57    | 45.6          | 698    | 1  | YMC_AECOL   | HYPOTHETICAL 78.7 KD L         |
| 14         | 57    | 45.6          | 713    | 1  | PALY_DICLA  | PHENYLALANINE AMMONIA          |
| 15         | 57    | 45.6          | 1116   | 1  | MHK1_SCHPO  | SERINE/THREONINE-PROTE         |
| 16         | 57    | 45.6          | 1314   | 1  | TEIX_CLOTE  | TETANUS TOXIN PRECURSO         |
| 17         | 56    | 44.8          | 1365   | 1  | KRES_YEAS   | KILLER TOXIN-RESISTANC         |
| 18         | 56    | 44.8          | 3555   | 1  | YAKB_YEAS   | HYPOTHETICAL 420.8 KD          |
| 19         | 55    | 44.0          | 260    | 1  | FLGG_SALTY  | FLAGELLAR BASAL-BODY R         |
| 20         | 55    | 44.0          | 294    | 1  | HMT_RAT     | HISTAMINE N-METHYLTRAN         |
| 21         | 55    | 44.0          | 378    | 1  | KLP2_BOMO   | KINESIN-LIKE PROTEIN K         |
| 22         | 55    | 44.0          | 482    | 1  | NIFO_AZOBR  | NITROGENASE MOLYBDENUM         |
| 23         | 55    | 44.0          | 491    | 1  | YF86_METJA  | HYPOTHETICAL PROTEIN M         |

| ID | SBP_CRYJA | STANDARD | PRO  | 374 AA |
|----|-----------|----------|------|--------|
| 24 | 54        | 43.2     | 159  | 1      |
| 25 | 54        | 43.2     | 292  | 1      |
| 26 | 54        | 43.2     | 298  | 1      |
| 27 | 54        | 43.2     | 315  | 1      |
| 28 | 54        | 43.2     | 331  | 1      |
| 29 | 54        | 43.2     | 334  | 1      |
| 30 | 54        | 43.2     | 334  | 1      |
| 31 | 54        | 43.2     | 337  | 1      |
| 32 | 54        | 43.2     | 483  | 1      |
| 33 | 54        | 43.2     | 492  | 1      |
| 34 | 54        | 43.2     | 621  | 1      |
| 35 | 54        | 43.2     | 702  | 1      |
| 36 | 54        | 43.2     | 737  | 1      |
| 37 | 54        | 43.2     | 743  | 1      |
| 38 | 54        | 43.2     | 774  | 1      |
| 39 | 54        | 43.2     | 1942 | 1      |
| 40 | 53        | 42.4     | 119  | 1      |
| 41 | 53        | 42.4     | 152  | 1      |
| 42 | 53        | 42.4     | 251  | 1      |
| 43 | 53        | 42.4     | 540  | 1      |
| 44 | 53        | 42.4     | 855  | 1      |
| 45 | 53        | 42.4     | 3343 | 1      |

ALIGNMENTS

RESULT

ID SBP\_CRYJA STANDARD PRO 374 AA

AC P18632:

DT 01-NOV-1990 (Rel. 16, Created)

DT 01-NOV-1995 (Rel. 32, Last sequence update)

DT 15-DEC-1998 (Rel. 37, Last annotation update)

DE SUGI BASIC PROTEIN PRECURSOR (SBP) (MAJOR ALLERGEN CRY J 1) (CRY J 1)

OS Cryptomeria japonica (Japanese cedar)

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Euphyllophytes; Spermatophyta; Coniferopsida; Coniferales; Taxodiaceae; Cryptomeria

RC SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.

RP TISSUE=POLLEN:

RA Sone I., Komiyama N., Shimizu K., Sakurai T., Morikubo K., Kine K., et al. Cloning and sequencing of cDNA encoding the major allergen of Japanese cedar pollen. J. Biochem. Biophys. Res. Commun. 199, 119-124 (1994)

RP SEQUENCE FROM N.A.

RC TISSUE=POLLEN:

| RA | Namba M., Kurose M., Torioka K., Fukuda S., Kurimoto M. | Submitted (JUL-1994) to the EMBL/GenBank/DBJ databases. |
| RL | [3] |  |
| RL | SEQUENCE OF 22-41. |  |
| RP | TISSUE=POLLEN: |  |
| RX | MEDLINE: 89031257. |  |
| RA | Taniaki M., Ando S., Usui M., Kurimoto M., Sakaguchi M., Inouye S., Matsui T. |  |
| RT | "N-terminal amino acid sequence of a major allergen of Japanese cedar pollen (Cry j 1)". |  |
| RL | FEBS Lett. 239:329-332(1988). |  |
| RP | [4] |  |
| RP | CARBOHYDRATES. |  |
| RA | MEDLINE: 95003748. |  |
| RA | Hijikata A., Matsumoto I., Kojima K., Ogawa H. |  |
| RT | "Antigenicity of the oligosaccharide moiety of the Japanese cedar (Cryptomeria japonica) pollen allergen, Cry j1". |  |
| RL | Int. Arch. Allergy Immunol. 105:198-202(1994). |  |
| CC | !- PTM: CONTAINS FUCOSEXYLOSE-CONTAINING N-LINKED OLIGOSACCHARIDES. |  |
| CC | !- DISEASE: THIS PROTEIN IS THE MAJOR ALLERGEN OF JAPANESE CEDAR POLLEN, THE MOST COMMON POLLEN ALLERGEN IN JAPAN. |  |
| CC | !- MISCELLANEOUS: THE SEQUENCE OF CRY J 1 FORM A IS SHOWN HERE. FORM B DIFFERS IN SIX POSITIONS. |  |
| CC | !- SIMILARITY: BELONGS TO THE OLYSACCHARIDE LYASE FAMILY 1. |  |

CC AMB A 1/AMB A 11/CRY J : SUBFAMILY  
CC  
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation  
CC at the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: D26544; AAA05542.1; 1;  
CC EMBL: D26545; AAA05543.1; 1;  
CC EMBL: D34639; AAA07020.1; 1;  
CC PIR: A44773; A44773;  
CC PRAM: PF00544; pectylase; 1;  
CC PRINIS: PR00807; AMB-ALLERGEN;  
CC Allergen: Glycoprotein; Multigene family; Signal;  
CC SIGNAL 1 21  
CC FT CHAIN 22 374 SUGI BASIC PROTEIN.  
CC FT VARIAT 14 143 L -> F (IN CRY J 1-B).  
CC FT VARIAT 202 202 H -> Y (IN CRY J 1-B).  
CC FT VARIAT 221 221 S -> T (IN CRY J 1-B).  
CC FT VARIAT 221 221 L -> S (IN CRY J 1-B).  
CC FT VARIAT 358 358 Q -> H (IN CRY J 1-B).  
CC FT VARIAT 361 361 K -> Q (IN CRY J 1-B).  
CC FT CARBOHYD 154 158 POTENTIAL.  
CC FT CARBOHYD 191 191 POTENTIAL.  
CC FT CARBOHYD 293 293 POTENTIAL.  
CC SEQUENCE 374 AA; 40645 MW; 74AB25950248F56F CRC64;  
Query Match: 74.48; Score 93; DR 1; Length 374;  
Best Local Similarity 70.0%; Pred. No. 4.05e-08;  
Matches 14; Conservative 3; Mismatches 3; Indels 0; Gaps 0;  
ID 312 SSGKVEGNIYIKKEAFNVE 351  
QY 1 SSGKVEGNIYIKKEAFNVE 20  
RESULT 2  
ID TKI HUMAN STANDARD: PRI: 623 AA.  
AC P29410:  
DI 01-DEC-1992 (Rel. 24, Created)  
DI 01-OCT-1996 (Rel. 34, Last sequence update)  
DI 01-OCT-1996 (Rel. 34, Last annotation update)  
DE TRANSKETOLASE (EC 2.2.1.1) (TK).  
GN TKTL OR TKL  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.  
RN [1]  
RS SEQUENCE FROM N.A.  
RX TISSUE=LIVER;  
RX MEDLINE: 93123263.  
RA McCool B.A., Pionk S.G., Martin P.R., Singleton C.K.;  
RT "Cloning of human transketolase cDNAs and comparison of the  
RT nucleotide sequence of the coding region in Wernicke-Korsakoff and  
RT non-Wernicke-Korsakoff individuals";  
RL J. Biol. Chem. 268:1397-1404(1993).  
RN [2]  
RS SEQUENCE FROM N.A.  
RX TISSUE=BRAIN;  
RX Crag F.H., Shew K.F.R.E., Szabo P., Plass J.P.;  
RA Submitted (SEP-1993) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RS SEQUENCE FROM N.A.  
RX TISSUE=BRAIN;  
RX Scherk G., Layfield R., Candy J.M., Duggleby R.G., Nixon P.F.;  
RA Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RS SEQUENCE OF 224-623 FROM N.A., AND PARTIAL SEQUENCE.  
RX MEDLINE: 92231878.

RA Abedinia M., Layfield R., Jones S.M., Nixon P.F., Mattick J.S.;  
RT "Nucleotide and predicted amino acid sequence of a cDNA clone  
RT encoding part of human transketolase";  
RL Biochem. Biophys. Res. Commun. 193:1159-1166(1992)  
CC 3-ALYLYTIC ACTIVITY: SEDOHEPTULOSE 7-PHOSPHATE - D-GLYCERALDEHYDE  
CC 3-PHOSPHATE - D-RIBOSE 5-PHOSPHATE - D-XULOSE 5-PHOSPHATE  
CC COFACTOR: THIAMINE PYOPHOSPHATE  
CC SUBUNIT: HOMODIMER.  
CC DISEASE: HAS BEEN IMPLICATED IN THE LATENT GENETIC DISEASE  
CC WERNICKE-KORSAKOFF SYNDROME (WKS) WHICH CAUSE SPECIFIC BRAIN  
CC DAMAGE.  
CC SIMILARITY: BELONGS TO THE TRANSKETOLASE FAMILY.  
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation  
CC at the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC EMBL: X67588; CAA47919.1; 1;  
CC EMBL: L22713; AAA61222.1; 1;  
CC EMBL: U55017; AAA98961.1; 1;  
CC PIR: A45050; A45050;  
CC PIR: PH0845; PH0845;  
CC MIM: 277730;  
CC PROSITE: PS00801; TRANSKETOLASE; 1;  
CC PROSITE: PS00802; TRANSKETOLASE; 1;  
CC PRAM: PF00456; transketolase; 1;  
KW Transferrase; Thiamine pyrophosphate;  
FT CONFLICT 30 31 IT -> SS (IN REF. 1);  
FT CONFLICT 45 46 I -> V (IN REF. 1);  
FT CONFLICT 224 230 LKAFGG -> AVQGLCE (IN REF. 4);  
FT CONFLICT 426 426 P A (IN REF. 1);  
FT CONFLICT 585 587 TML -> KTM (IN REF. 2);  
FT CONFLICT 608 623 LKLAQAVRGLTITKA -> TQPLHLK (IN REF. 4);  
FT SEQUENCE 623 AA; 67877 MW; 176B9032F627125 CRC64;  
Query Match: 54.48; Score 88; DR 1; Length 623;  
Best Local Similarity 50.0%; Pred. No. 2.31e-02;  
Matches 5; Conservative 4; Mismatches 5; Indels 0; Gaps 0;  
ID 472 TSPPENAIYNNNEAFV 489  
QY 2 TSPPENAIYNNNEAFV 24  
RESULT 3  
ID TKI MOUSE STANDARD: PRI: 623 AA.  
AC P40142:  
DI 01-FEB-1995 (Rel. 31, Created)  
DI 01-FEB-1995 (Rel. 31, Last sequence update)  
DI 15-FEB-2000 (Rel. 39, Last annotation update)  
DE TRANSKETOLASE (EC 2.2.1.1) (TK) (P66).  
GN TKL  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
RN [1]  
RS SEQUENCE FROM N.A.  
RX STRAIN=LAF1;  
RX MEDLINE: 96214928.  
RA Schimmer B.P., Tsao J., Czerwinski W.;  
RT "Amplification of the transketolase gene in desensitization-resistant  
RT mutant mouse adrenocortical tumor cells";  
RL J. Biol. Chem. 271:4993-4998(1996).  
RN [2]  
RS SEQUENCE OF 1-36 FROM N.A.  
RX STRAIN=129/SV;  
RX Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.  
RA Salamon C., Chervenak M., Piatigorsky J., Sax C.M.;  
RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.  
CC 3-ALYLYTIC ACTIVITY: SEDOHEPTULOSE 7-PHOSPHATE - D-GLYCERALDEHYDE



```

Db 472 TSRENATYISNNEDQV 489
QY 2 SGRNEGNTYNNNEAFK 19

RESULT 5
ID SW16 YEAST STANDARD PRI 803 AA
AC P09959:
DT 01-MAR-1989 (Rel. 10, Created)
DI 01-MAR-1989 (Rel. 10, Last sequence update)
DE 01-OCT-1994 (Rel. 34, Last annotation update)
DT REGULATORY PROTEIN SW16 (CELL-CYCLE BOX FACTOR, CHAIN SW16) (TRANS-
DE ACTING ACTIVATOR OF HO ENDONUCLEASE GENE) (YBF SUBUNIT P90).
GN SW16 OR YLR182W OR Y470.8.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales.
OC Saccharomycetaceae; Saccharomycetes.
RN [...]
RP SEQUENCE FROM N.A.
RX MEDLINE: 86014241.
RA Breeden L., Nasmyth K.:
RI "Similarity between cell-cycle genes of budding yeast and fission
RI yeast and the Notch gene of Drosophila."
RI Nature 329:651-654(1987).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB572:
RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Du Z.,
RA Favell A., Fulton L., Gattung S., Greco T., Kirsten J.,
RA Kucaba T., Hallsworth K., Hawkins J., Hillier S., Jier M.,
RA Johnson D., Johnston I., Lanzaston Y., Laruelle P., Le
RA Mardis E., Menezes S., Miller N., Nhan M., Pauley A., Peluso D.,
RA Riken L., Riles L., Raich A., Trevisan E., Vignati D.,
RA Wilcox L., Wolfdmnan P., Vaudin M., Wilson K., Waterston R.
RL Submitted (NOV-1994) to the EMBL/GenBank/CDRJ databases.
RN [3]
RP CHARACTERIZATION.
RX MEDLINE: 94201535.
RA Dirlik U., Moll T., Auer H., Nasmyth K.:
RI "A central role for Swi6 in modulating cell cycle start-specific
RI transcription in yeast."
RI Nature 357:508-513(1992).
CC -- FUNCTION: PART OF A COMPLEX INVOLVED IN CELL-CYCLE-DEPENDENT
CC CELL-CYCRIPTION. SW14 AND SW16 ARE REQUIRED FOR FORMATION OF THE
CC CELL-CYCLE BOX FACTOR-DNA COMPLEX. THE REPEATED ELEMENT IN THE
CC UPSTREAM REGION OF HO (5'-GAGGAGAA-3') IS CALLED THE CELL-CYCLE
CC BOX (CCB).
CC -- SUBUNIT: YBF CONTAINS SW16 AND YBF1. SW16 CONTAINS SW16 AND SW14.
CC -- SIMILARITY: STRONG TO S-PERME CELL.
CC -- SIMILARITY: CONTAINS 2 ANK REPEATS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collabora-
CC between the Swiss Institute of Bioinformatics and the EMBL Outposts
CC the European Bioinformatics Institute. There are no restrictions
CC use by non-profit institutions as long as its content is in
CC modified and this statement is not removed. Usage by and for com-
CC entities requires a license agreement (See http://www.isb-sib.ch/an
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: X05238; CAA29581.1;
CC EMBL: U17246; AAB67460.1;
CC PIR: S03161; RCBYW6.
CC
CC TRANSFAC: T00096;
CC TRANSFAC: T01013;
CC SGD: L0002254; SW16.
CC
CC PTAM: PF00023; ank: 2.
CC
CC TRANSCRIPTION regulation: DNA-binding; ANK repeat; Repeat.
CC
CC FT DOMAIN 311 495 2 X ANK MOTIF REPEATS.
CC REPEAT 311 343 ANK MOTIF 1.
CC FT REPEAT 453 495 ANK MOTIF 2.
CC FT DOMAIN 631 640 GLU-RICH (ACIDIC).
CC
CC SEQUENCE 803 AA; 90559 MW; 9B317FCACFE9493C CRC64:

```

\*P433, a Saccharomyces cerevisiae gene encoding a peroxisomal integral membrane protein essential for peroxisome biogenesis". J. Cell Biol. 114:1167-1178(1991).

[2]

SEQUENCE FROM N.A.

STRAN=5288C / AB972;

Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Li Z., Favello A., Fulton L., Gallung S., Gress T., Kirsten J., Kucaba D., Hallsworth K., Hawkins J., Hillier J., Jier M., Johnson D., Johnston L., Langsten Y., Catreille P., Le I., Mardis E., Meneses S., Miller N., Nhan M., Pauley A., Peluso D., Rifkin L., Riley L., Taich F., Trevasaki E., Vignati D., Wilcox L., Wolhdman P., Vauc M., Wilson R., Waterston R.; Submitted (Jul-1995) to the EMBL/GenBank/DDBJ databases.

- FUNCTION: INVOLVED IN PEROXISOME BIOSYNTHESIS; MAY FUNCTION AS A RECEPTOR PROTEIN. LACK OF THIS PROTEIN CAUSES THE PEROXISOMAL-DEFICIENT PHENOTYPE AND MISLOCALIZATION IN THE CYTOSOL OF PEROXISOMAL MATRIX PROTEINS.

- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. PEROXISOMAL.

- INDUCTION: BY OLEIC ACID.

-----

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sit.ch/announce/> or send an email to [license@isb.ch](mailto:license@isb.ch))

-----

EMBL: X58407; CAA43309.1;  
EMBL: U32517; AAB64764.1;  
PIR: A40550; A40550.  
SGD: L0001339; PAS3.  
Transmembrane: Peroxisome.  
DOMAIN 1 17 MATRIX (POTENTIAL).  
TRANSFAM 18 39 POTENTIAL.  
DOMAIN 40 441 CYTOPLASMIC (POTENTIAL).  
SEQUENCE 441 AA; 55675 MW; 3CD51G336FE0DF9 CRC64;

Query Match 47.2%; Score 59; DB 1; Length 441;  
Best Local Similarity 45.7%; Pred No. 1,56e+00;  
Matches 7; Conservative 6; Mismatches 2; Indels 0; Gaps 0;

Ddb 212 SKKEQGVYNEQAF 226  
 : ||| ||| ||| |  
 : GKNEGINYYNNEAF 17

RESULT 8 STANDARD: PST: 80 AA.

MILD YES-HAELIN  
AC P442C4:  
GI-NOV-1995 (Rel. 32, Created)  
CI-NOV-1995 (Rel. 32, Last sequence update)  
OI-NOV-1995 (Rel. 32, Last annotation update)  
DE HYPOTHEITICAL PROTEIN HI1458.  
GN HI1458  
OS Haemophilus influenzae.  
OC Bacteria; Proteobacteria; gamma subdivision: Pasteurellaceae;  
CC Haemophilus.  
RN [1]  
RS SEQUENCE FROM N.A.  
RC STRAIN-RD / KW20;  
RX MEDLINE: 9530630.  
RA Fleischmann F.D., Adams M.D., White O., Clayton R.A., Kirkness E.F., Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M., McInerney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D., Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M., Weidma J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.B., Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.W., Brandon R.C., Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghegan N.S.M., Gnehm C.S., McDonald L.A., Small K.V., Fraser C.M., Smith H.O., Venter J.C.;  
RT "Whole-genome random sequencing and assembly of Haemophilus

RI Influenzae Rd.":  
 RL Science 269:456-512(1995).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: J02824; AAC23111.1; .  
 CC TIGR: H1458; .  
 KW Hypothetical protein.  
 SQ SEQUENCE 80 AA: 9400 MW: E430D2FC7714CDD CRC64:  
 Query Match 45.48: Score 58: DB 1: Length 80:  
 Best Local Similarity 42.18: Pred. No. 2.43e-00:  
 Matches 8: Conservative 7: Mismatches 4: Indels 0: Gaps 0:  
 DE 28 NGNNTSYVGNMESLIE 46  
 CY 2 SGKKEGINNNEAFKVE 20  
 CC -----  
 CC RESULT 9  
 CC ID TTKB-DROME STANDARD: PRI: 641 AA.  
 CC AC P17789;  
 CC DT 01-AUG-1990 (Rel. 15, Created)  
 CC DT 01-AUG-1990 (Rel. 15, Last sequence update)  
 CC DT 15-JUL-1998 (Rel. 36, Last annotation update)  
 CC DE TRANSCRIPT PROTEIN, BETA ISOFORM (TRAMTRA / P69) (YUSHI TARAZU  
 CC DE REPRESSOR PROTEIN).  
 CC GN TTK OR FTZF2/HH.  
 CC OS Drosophila melanogaster (fruit fly).  
 CC OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 CC OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 CC OC Rhytridae; Drosophilidae; Drosophila.  
 CC [1]  
 CC SEQUENCE FROM N.A.  
 CC MEDLINE: 90107945.  
 CC RA Harrison S.D., Travers A.A.;  
 CC "The tramtrack gene encodes a Drosophila finger protein that  
 CC interacts with the ftz transcriptional regulatory region and shows a  
 CC novel embryonic expression pattern.";  
 CC EMBO J. 9:207-216(1990).  
 CC [2]  
 CC SEQUENCE FROM N.A.  
 CC MEDLINE: 90107945.  
 CC RA Brown J.L., Sonoda S., Ueda H., Scott M.P., Wu C.;  
 CC Submitted (XX-1991) to the EMBL/GenBank/DBJ databases.  
 CC [3]  
 CC SEQUENCE FROM N.A.  
 CC MEDLINE: 93279467.  
 CC RA Xiong W.C., Montell C.;  
 CC "Tramtrack is a transcriptional repressor required for cell fate  
 CC determination in the Drosophila eye.";  
 CC Genes Dev. 7:1085-1095(1993).  
 CC [4]  
 CC DNA-BINDING STUDIES.  
 CC MEDLINE: 92349422.  
 CC RA Fairall L., Harrison S.D., Travers A.A., Rhodes D.;  
 CC "Sequence-specific DNA binding by a two zinc-finger peptide from the  
 CC Drosophila melanogaster tramtrack protein.";  
 CC J. Mol. Biol. 226:349-366(1992).  
 CC [5]  
 CC X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS) OF 499-561.  
 CC MEDLINE: 94057350.  
 CC RA Fairall L., Schwabe J.W.R., Chapman L., Finch J.T., Rhodes D.;  
 CC "The crystal structure of a two zinc-finger peptide reveals an  
 CC extension to the rules for zinc-finger/DNA recognition.";  
 CC Nature 366:483-487(1993).  
 CC RL Nature 366:483-487(1993).  
 CC !- FUNCTION: BINDS TO A NUMBER OF SITES IN THE TRANSCRIPTIONAL  
 CC REGULATORY REGION OF FTZ. MAY FUNCTION BY REPRESSING INAPPROPRIATE

CC SEGMENTATION GENE TRANSFORMATION DURING EMBRYOGENESIS. P69 IS  
 CC REQUIRED TO REPRESS THE L1 PROMOTION OF GENES THAT ARE INCOMPATIBLE  
 CC WITH THE DEVELOPMENT OF PHOTORECEPTOR CELL FATES. MAY BIND TO  
 CC THE REGION 5'-GGTCTCTCC-3'.  
 CC !- SUBUNIT: MONOMER.  
 CC !- SUBCELLULAR LOCATION: NUCLEAR.  
 CC !- ALTERNATIVE PRODUCTS: TWO ISOFORMS, ALPHA AND BETA, ARE PRODUCED  
 CC BY ALTERNATIVE SPLICING OF THE C-TERMINAL DOMAIN.  
 CC !- DEVELOPMENTAL STAGE: EXPRESSED THROUGHOUT EMBRYOGENESIS. DURING  
 CC THE FAST PHASE OF GERM AN ELONGATION, THIS FORM IS PRESENT IN  
 CC YOLK NUCLEI.  
 CC !- SIMILARITY: CONTAINS A TERMINAL BR-C/TTK DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: X17121; CAA34981.1; .  
 CC EMBL: M62856; AAA28544.1; .  
 CC EMBL: X71627; CAA50634.1; ALT\_INIT.  
 CC PIR: S10881; S10881.  
 CC PDB: 2DRP; 31-AUG-94.  
 CC TRANSFAC: T00843; .  
 CC FLYBASE: FBgn003870; ttk.  
 CC PROSITE: PS00029; ZINC\_FINGER\_C2H2\_2.  
 CC PFAM: PF00096; zf-C2H2\_2.  
 CC PFAM: PF00681; ZTB\_1.  
 CC KW Transcription regulation; Zinc-finger; Metal-binding; DNA-binding;  
 CC Nuclear protein; 3D-structure; Alternative splicing.  
 CC FT DOMAIN 1 114 BR-C/TTK DOMAIN.  
 CC FT ZN\_FING 506 529 C2H2-TYPE.  
 CC FT ZN\_FING 536 559 C2H2-TYPE.  
 CC FT CONFLICT 253 253 T -> M (IN REF. 2).  
 CC FT CONFLICT 279 279 H -> Q (IN REF. 2).  
 CC FT CONFLICT 499 499 -> I (IN REF. 2).  
 CC SQ SEQUENCE 641 AA: 6951 MW: B5EB54128171688 CRC64:  
 Query Match 45.48: Score 58: DB 1: Length 641:  
 Best Local Similarity 54.58: Pred. No. 2.43e-00:  
 Matches 6: Conservative 1: Mismatches 3: Indels 0: Gaps 0:  
 DB 444 DLYNSNDGFKL 454  
 CY 9 NLYNNNAFKV 19  
 CC -----  
 CC RESULT 10  
 CC ID FLGG-BORBU STANDARD: PRI: 265 AA.  
 CC AC C51715;  
 CC DT 15-DEC-1998 (Rel. 37, Created)  
 CC DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 CC DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 CC DE FLAGELLAR BASAL-BODY ROD PROTEIN FLGG (DISTAL ROD PROTEIN).  
 CC GN FLGG OR BB0774.  
 CC OS Borrelia burgdorferi (Lyme disease spirochete).  
 CC OC Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.  
 CC [1]  
 CC SEQUENCE FROM N.A.  
 CC STRAIN-ATCC 35210 / B31;  
 CC MEDLINE: 98055943.  
 CC RA Fraser C.M., Casjers S., Huar W.M., Sutton G.G., Clayton R.A.,  
 CC Lathigra R., White O., Ketchu K.A., Dodson R., Hickey E.K., Gwinn M.,  
 CC Dougherty B., Tomb J.F., Fleischmann R.D., Richardson D.,  
 CC Persker J., Kierulff A.R., Quackenbush J., Salzberg S., Hanson M.,  
 CC Van Vagt R., Palmer N., Adams M.D., Gocayne J.D., Weidman J.,  
 CC Utterback T., Wathley L., McDonald L., Artlich P., Bowman C.,  
 CC Garland S., Fujii C., Cotton M.D., Horst K., Roberts K., Hatch R.,  
 CC Smith H.O., Venter J.C.;  
 CC "Genomic sequence of a Lyme disease spirochete, Borrelia



[illegible]

```

RN      (2)
RP      SEQUENCE FROM N.A
RC      STRAIN-972:
RA      Connor R., Church C.M., Barrell B.G., Ralandream M.A., Walsh S.V.:
RL      Submitted (Apr-1996) to the EMBL/GenBank/DBJ databases.
CC      -!- FUNCTION: MAY REGULATE CELL MORPHOLOGY, CELL WALL INTEGRITY, SALT
CC      RESISTANCE, CELL CYCLE REENTRY FROM STATIONARY-PHASE ARREST, AND
CC      FILAMENTOUS GROWTH IN RESPONSE TO STRESS.
CC      -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC      MAP KINASE KINASE SUBFAMILY.
CC      -----
CC      THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL: U53872; AAB62119.1;
DR      EMBL: Z70690; CAA94523.1;
DR      HSSP: P24941; IAC1.
DR      PROSITE: PS00107; PROTEIN_KINASE_ATP_1;
DR      PROSITE: PS00108; PROTEIN_KINASE_ST_1;
DR      PROSITE: PS50011; PROTEIN_KINASE_DOM_1;
DR      PFAM: PF00069; pkinase; 1.
KW      Transferase; Serine/threonine-protein kinase; Tyrosine-protein kinase;
KW      ATP-binding.
FT      DOMAIN      825 1094      PROTEIN KINASE.
FT      NP_BIND      831 839      ATP (BY SIMILARITY).
FT      BINDING      854 854      ATP (BY SIMILARITY).
FT      ACT_SITE      955 955      BY SIMILARITY.
SQ      SEQUENCE 1116 AA; 225132 MW; 7AFDB9EC62ED47B CRC64;
Query Match 45.64; Score 57; DB 1; Length 1116;
Best Local Similarity 40.04; Pred. No. 3.76e+00;
Matches 6; Conservative 4; Mismatches 5; Indels 0; Gaps 0;
Db 402 KADDSIYHKEDEK 416
vY 4 KNEGNIYNNNEAFK 18

```

Search completed: Mon Jun 19 16:14:43 2000  
 Job time : 7 secs.

\*\*\*\*\*

WATERMAN

(TM)

\*\*\*\*\*

Release 3.1A John F. Collins, BioComputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Mon Jun 19 16:14:59 2000: Maspar time 9.10 Seconds  
152 449 Million cell updates/sec  
Isolular output not generated.

Title: >US-09-142-524A-12  
Description: (1-25) from US9142524A.pep  
Perfect Score: 125  
Sequence: 1 SSGKNEGNIYNNNEAFKVE 20

Scoring table: PAM 150  
Gap 15

Searched: 225878 seqs. 69334122 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: splrembl12  
1:sp-archaea 2:sp-bacteria 3:sp-fungi 4:sp-human  
5:sp-invertebrate 6:sp-mammal 7:sp-mtc 8:sp-organelle  
9:sp-phage 10:sp-plant 11:sp-rodent 12:sp-unclassified  
13:sp-vertebrate 14:sp-virus

Statistics: Mean 26.556; Variance 35.964; scale 0.738

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description                    | Pred. No. |
|------------|-------|-------------|--------|-------|--------------------------------|-----------|
| 1          | 125   | 100.0       | 375    | 10    | Q95385 CHAO1 PRECURSOR.        | 2.25e-15  |
| 2          | 108   | 86.4        | 367    | 10    | Q92NU7 POLLEN MAJOR ALLERGEN   | 4.37e-11  |
| 3          | 69    | 55.2        | 1598   | 10    | Q92SS7 DYNEIN HEAVY CHAIN ISO  | 4.01e-02  |
| 4          | 53    | 50.4        | 397    | 3     | G08144 CHROMOSOME XV READING   | 6.52e-01  |
| 5          | 51    | 48.8        | 420    | 2     | Q92H17 NADH OXIDASE (FRAGMENT  | 1.60e-00  |
| 6          | 60    | 48.0        | 126    | 13    | Q73615 TRANSEKTOXASE (FRAGMENT | 2.48e+00  |
| 7          | 59    | 47.2        | 146    | 2     | Q45841 17 KD HEMAGGLUTININ CO  | 3.84e+00  |
| 8          | 59    | 47.2        | 146    | 2     | Q45878 HA17.                   | 3.84e+00  |
| 9          | 59    | 47.2        | 264    | 2     | O83926 FLAGELLAR BASAL-BODY R  | 3.84e+00  |
| 10         | 59    | 47.2        | 298    | 5     | O19058 SIMILAR TO HYDRATASE-D  | 3.84e+00  |
| 11         | 58    | 46.4        | 290    | 2     | Q59078 BETA-ACARASE I PRECURS  | 5.92e+00  |
| 12         | 58    | 46.4        | 379    | 14    | Q9W442 ORF MSV050 HYPOTHETICAL | 5.92e+00  |
| 13         | 58    | 46.4        | 432    | 2     | Q97980 HYPOTHETICAL 50.7 KD P  | 5.92e+00  |
| 14         | 57    | 45.6        | 2206   | 5     | O95205 HYPOTHETICAL 264.1 KD   | 9.07e+00  |
| 15         | 57    | 45.6        | 4550   | 5     | O77336 PFC0425W PROTEIN.       | 9.07e+00  |
| 16         | 57    | 45.6        | 4981   | 5     | O77372 PFC0820W PROTEIN.       | 9.07e+00  |
| 17         | 56    | 44.8        | 163    | 2     | O01093 SMFE PROTEIN PRECURSOR  | 1.38e+01  |
| 18         | 56    | 44.8        | 321    | 9     | Q92XD3 ORF50.                  | 1.38e+01  |
| 19         | 56    | 44.8        | 334    | 2     | Q92168 1, 2-DIHYDROXYBENZYL-PY | 1.38e+01  |
| 20         | 55    | 44.8        | 383    | 2     | Q92N11 LYTN.                   | 1.38e+01  |

|    |    |      |      |    |                               |          |
|----|----|------|------|----|-------------------------------|----------|
| 21 | 56 | 44.8 | 416  | 2  | G92HJ3 NADH OXIDASE (FRAGMENT | 1.38e+01 |
| 22 | 56 | 44.8 | 423  | 2  | G92H12 NADH OXIDASE (FRAGMENT | 1.38e+01 |
| 23 | 56 | 44.8 | 424  | 2  | G92H19 NADH OXIDASE (FRAGMENT | 1.38e+01 |
| 24 | 56 | 44.8 | 486  | 2  | Q2G3P2 OLIGOPEPTIDE TRANSPORT | 1.38e+01 |
| 25 | 56 | 44.8 | 929  | 2  | G9XJ77 HYPOTHETICAL 101.7 KD  | 1.38e+01 |
| 26 | 55 | 44.0 | 204  | 4  | G9Y555 BK150C2.3 (PUTATIVE N  | 2.10e+01 |
| 27 | 55 | 44.0 | 413  | 14 | C41627 ENVELOPE GLYCOPROTEIN  | 2.10e+01 |
| 28 | 55 | 44.0 | 430  | 5  | G9XXV2 MAPK-RELATED KINASE    | 2.10e+01 |
| 29 | 55 | 44.0 | 524  | 1  | O59209 514AA LONG HYPOTHETICA | 2.10e+01 |
| 30 | 55 | 44.0 | 608  | 5  | O96849 SLIME MOLD (DICTYOSTE  | 2.10e+01 |
| 31 | 55 | 44.0 | 688  | 5  | 23898 HYPOTHETICAL 70.4 KO P  | 2.10e+01 |
| 32 | 55 | 44.0 | 888  | 2  | 15554 101 KDA PROTEIN PRECUR  | 2.10e+01 |
| 33 | 55 | 44.0 | 1039 | 1  | G9YF27 1039AA LONG HYPOTHETIC | 2.10e+01 |
| 34 | 55 | 44.0 | 1038 | 5  | C77306 PFC0105W PROTEIN.      | 2.10e+01 |
| 35 | 55 | 44.0 | 3078 | 5  | O26031 VARIANT-SPECIFIC SUREA | 2.10e+01 |
| 36 | 55 | 44.0 | 3844 | 5  | O94646 AARL1 PROTEIN (FRAGMEN | 2.10e+01 |
| 37 | 54 | 43.2 | 330  | 2  | Q92H66 HYDRATASE/ALDOLASE PHN | 3.16e+01 |
| 38 | 54 | 43.2 | 337  | 2  | Q928F3 HOLLYDAY JUNCTION HELI | 3.16e+01 |
| 39 | 54 | 43.2 | 1064 | 10 | C04884 LYSINE-KETOGLUTARATE R | 3.16e+01 |
| 40 | 54 | 43.2 | 1064 | 10 | C04884 LYSINE-KETOGLUTARATE R | 3.16e+01 |
| 41 | 54 | 43.2 | 1184 | 5  | O7762 MAL3P4.27 PROTEIN       | 3.16e+01 |
| 42 | 54 | 43.2 | 1200 | 5  | O92777 PFC0860W PROTEIN       | 3.16e+01 |
| 43 | 54 | 43.2 | 1306 | 5  | O96129 PREDICTED MEMBRANE ASS | 3.16e+01 |
| 44 | 54 | 43.2 | 2708 | 5  | O15791 CHLOROQUINE RESISTANCE | 3.16e+01 |
| 45 | 54 | 43.2 | 2815 | 5  | O15792 STRAIN HB3 CG2 (CG2).  | 3.16e+01 |

ALL: NMENTS

|   |                                   |                     |             |
|---|-----------------------------------|---------------------|-------------|
| RESULT 1  | PRELIMINARY:                      | PRT:                | 375 AA.     |
| ID Q96395   |                                   |                     |             |
| AC Q96395   |                                   |                     |             |
| DI 01-FEB-1997 (TRENBLrel. 02, Created)                                     |                                   |                     |             |
| DT 01-FEB-1997 (TRENBLrel. 02, Last sequence update)                        |                                   |                     |             |
| DI 01-NOV-1999 (TRENBLrel. 12, Last annotation update)                      |                                   |                     |             |
| DE CHAO1 PRECURSOR.   |                                   |                     |             |
| OS Chamaecyparis obtusa.  |                                   |                     |             |
| OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;       |                                   |                     |             |
| OC euphyllophytes; Spermatophyta; Coniferopsida; Coniferales;               |                                   |                     |             |
| OC Taxodiaceae; Chamaecyparis.  |                                   |                     |             |
| RN [1]  |                                   |                     |             |
| RP SEQUENCE FROM N.A.   |                                   |                     |             |
| RC TISSUE-POLLEN:   |                                   |                     |             |
| RX MEDLINE: 96265194.   |                                   |                     |             |
| RA SUZUKI M., KOMIYAMA N., ITOH M., ITOH H., SONE T., KUNIKI K., IAKARI I., |                                   |                     |             |
| RA CHIA N.  |                                   |                     |             |
| RT "Purification, characterization and molecular cloning of Cha o 1, a      |                                   |                     |             |
| RI major allergen of Chamaecyparis obtusa (Japanese cypress) pollen"        |                                   |                     |             |
| RL Vol. Immuni.: 33:451-460(1996).  |                                   |                     |             |
| DR EMBL: D45404; BAA08246.1; ..   |                                   |                     |             |
| DR MENDEL: 7526; Chao01:1088:7626.  |                                   |                     |             |
| DR PFAM: PF00544; pec_lyase: 1.   |                                   |                     |             |
| DR PRINTS: PR00807; AMBALLERGEN.  |                                   |                     |             |
| KW Signal.  |                                   |                     |             |
| FT SIGNAL.  | 1                                 | 2:                  | POTENTIAL.  |
| FT CHAIN.   | 22                                | 375                 | CHAO1.      |
| SQ SEQUENCE   | 375 AA: 40258 MW: A0981492 CRC32: |                     |             |
| Query Match   | 100.0%:                           | Score 125; DB 10;   | Length 375; |
| Best Local Similarity   | 100.0%:                           | Pred. No. 2.25e-15; |             |
| Matches   | 20: Conservative                  | 0: Mismatches       | 0: Indels   |
|   | 0: Gaps                           | 0:                  |             |
| Db  | 332 SSGKNEGNIYNNNEAFKVE 351       |                     |             |
| QY  | 1 SSGKNEGNIYNNNEAFKVE 20          |                     |             |
| RESULT 2  | PRELIMINARY:                      | PRT:                | 367 AA.     |
| ID Q92NU7   |                                   |                     |             |
| AC Q92NU7   |                                   |                     |             |
| DI 01-MAY-1999 (TRENBLrel. 10, Created)                                     |                                   |                     |             |
| DT 01-MAY-1999 (TRENBLrel. 10, Last sequence update)                        |                                   |                     |             |
| DI 01-NOV-1999 (TRENBLrel. 12, Last annotation update)                      |                                   |                     |             |

DB POLLEN MAJOR ALLERGEN 1-2.  
 OS Juniperus ashei (Ozark white cedar).  
 OC Eukaryota: Viridiplantae; Streptophyta; Tracheophyta;  
 OC euphyllophytes; Spermatophyta; Coniferopsida; Coniferales;  
 OC Taxodiaceae; Juniperus.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MIDORE-HORIUTI T.M., GOLDELM R.M., KUROSKI A., WOEI T.G.,  
 RA BROOKS E.G.,  
 RI "Molecular cloning of mountain cedar (Juniperus ashei) pollen major  
 RI allergen, Jun 4 1999;  
 RC Submitted (NCV-1996) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF106663; AAC03609; ...  
 DR EMBL: AF106662; AAC03608; ...  
 DR MENDEL: 36544; Lucas:1088;15544.  
 DR MENDEL: 36545; Lucas:1088;36545.  
 SQ SEQUENCE 367 AA: 39824 MW: 4C2D5630 CRC32:  
 Query Match 86.4% Score 108; DB 10; Length 367;  
 Best Local Similarity 85.0% Pred. No. 4 37e-11;  
 Matches 17; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
 DB 332 SSGTEEN:YNSNEAKVE 35;  
 QY 1 SSGKNEGTINYNNEAKVE 20  
 RESULT 4  
 ID Q9ZSS7 PRELIMINARY: PRT: 1198 AA.  
 AC Q9ZSS7  
 DT 01-MAY-1999 (TrEMBLrel: 10, Created)  
 DI 01-MAY-1999 (TrEMBLrel: 10, Last sequence update)  
 DE 01-MAY-1999 (TrEMBLrel: 10, Last annotation update)  
 DE DYNEIN HEAVY CHAIN ISOFORM DHC1B (FRAGMENT).  
 RN DHC1B.  
 OS Chlamydomonas reinhardtii.  
 OC Eukaryota: Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;  
 OC Chlamydomonadales; Chlamydomonadales.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA PAZOUK G.J., DICKERT B.L., WITMAN G.B.,  
 RI "The dynein heavy chain isoform DHC1B is required for flagellar  
 RI assembly".  
 RC Submitted (Oct 1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF096277; AAC9457.1;  
 FT NON-TER 1  
 FT NON-TER 1198  
 SQ SEQUENCE 1198 AA: 334914 MW: 45558426 CRC32:  
 Query Match 55.2% Score 69; DB 10; Length 1198;  
 Best Local Similarity 35.3% Pred. No. 4 01e-02;  
 Matches 6; Conservative 10; Mismatches 1; Indels 0; Gaps 0;  
 DB 159 REESAKYKEAEAKID 175  
 QY 4 KNEGTINYNNEAKVE 20  
 RESULT 4  
 ID Q08144 PRELIMINARY: PRT: 397 AA.  
 AC Q08144  
 DT 01-NOV-1996 (TrEMBLrel: 01, Created)  
 DI 01-NOV-1996 (TrEMBLrel: 01, Last sequence update)  
 DE 01-NOV-1996 (TrEMBLrel: 08, Last annotation update)  
 DE CHROMOSOME XV READING FRAME ORF Y01018C.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;  
 OC Saccharomycetaceae; Saccharomycetes.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA HUGHES B., POHL T.M.,  
 RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.  
 RN [2]

RP SEQUENCE FROM N.A.  
 RA MIPS;  
 RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: Z74760; CAA95017.1;  
 DR PFAM: PF00894; Syntaxin1.1  
 SQ SEQUENCE 397 AA: 45875 MW: 79FBIAY9 CRC32:  
 Query Match 50.4% Score 63; DB 3; Length 397;  
 Best Local Similarity 51.3% Pred. No. 6 52e-01;  
 Matches 8; Conservative 6; Mismatches 1; Indels 0; Gaps 0;  
 DB 342 SSGRNGSNKYNND 356  
 QY 1 SSGKNEGTINYNNE 15  
 RESULT 5  
 ID Q9ZHT7 PRELIMINARY: PRT: 420 AA.  
 AC Q9ZHT7  
 DT 01-MAY-1999 (TrEMBLrel: 10, Created)  
 DI 01-MAY-1999 (TrEMBLrel: 10, Last sequence update)  
 DE 01-NOV-1999 (TrEMBLrel: 12, Last annotation update)  
 DE NADH OXIDASE (FRAGMENT).  
 OS Serpulina sp. P280/1.  
 OC Bacteria; Spirochaetales; Spirochaetaceae; Serpulina.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA STRAIN-P280/1.  
 RA ATYEO R.F., JENSEN N.S., STANTON T.B., HAMPTON D.J.,  
 RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF060815; AAC78824.1;  
 DR HSP: P37062; LJOA.  
 FT NON-TER 1  
 FT NON-TER 420  
 SQ SEQUENCE 420 AA: 46594 MW: 1E88D66 CRC32:  
 Query Match 49.8% Score 61; DB 2; Length 420;  
 Best Local Similarity 35.3% Pred. No. 1 60e-00;  
 Matches 6; Conservative 1; Mismatches 6; Indels 0; Gaps 0;  
 DB 54 KGEIDVYMGHEVWKID 70  
 QY 4 KNEGTINYNNEAKVE 20  
 RESULT 6  
 ID Q73616 PRELIMINARY: PRT: 121 AA.  
 AC Q73616  
 DT 01-AUG-1998 (TrEMBLrel: 07, Created)  
 DI 01-AUG-1998 (TrEMBLrel: 07, Last sequence update)  
 DE 01-AUG-1998 (TrEMBLrel: 07, Last annotation update)  
 DE TRANSKETOLASE (FRAGMENT).  
 GN IK.  
 OS Xenopus laevis (African clawed frog).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Amphibia;  
 OC Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae; Xenopodinae;  
 OC Xenopus.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA SPINDLER B., VERREY F.,  
 RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AJ224484; CAA11971.1;  
 FT NON-TER 1  
 FT NON-TER 126  
 SQ SEQUENCE 126 AA: 13539 MW: B740600F CRC32:  
 Query Match 48.0% Score 60; DB 13; Length 126;  
 Best Local Similarity 33.3% Pred. No. 2 48e-00;  
 Matches 6; Conservative 7; Mismatches 5; Indels 0; Gaps 0;  
 DB 21 TSPEDAVIYSSTEEFKI 38  
 QY 1 SSGKNEGTINYNNEAKVE 20



QY 2 SSKNEGTNIYNNNEAFKV 19

RESULT 7  
 ID Q45841 PRELIMINARY: PRT: 146 AA.  
 AC Q45841  
 DT 01-NOV-1996 (TRENBLrel. 01, Created)  
 DT 01-NOV-1996 (TRENBLrel. 02, Last sequence update)  
 DE 01-NOV-1996 (TRENBLrel. 07, Last annotation update)  
 DE 17 KB HEMAGGLUTININ COMPONENT.  
 GN HEM17/B.  
 OS Clostridium botulinum.  
 OC Bacteria: Firmicutes; Bacillus/Clostridi m group; Clostridiaceae;  
 OC Clostridium.  
 RN [1]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN-LAMANNA:  
 RA YANG G., RHEE S., JUNG H., YANG K.  
 RI Submitted (APR-1995) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: U2443; AAA99056.1; .  
 SQ SEQUENCE 146 AA: 15908 MW: 6150A4C9 CRC32:

Query Match 47.2% Score 59: DB 2: Length 146;  
 Best Local Similarity 58.3%; Pred. No. 3.84e+00;  
 Matches 7: Conservative 4: Mismatches 1: Indels 0: Gaps 0:  
 Db 133 DIYNSQWFKLE 144  
 QY 9 NIYNNNEAFKVE 20

RESULT 8  
 ID Q45878 PRELIMINARY: PRT: 146 AA.  
 AC Q45878  
 DT 01-NOV-1996 (TRENBLrel. 01, Created)  
 DT 01-NOV-1996 (TRENBLrel. 02, Last sequence update)  
 DE 01-NOV-1996 (TRENBLrel. 10, Last annotation update)  
 DE HA17.  
 GN HA17 OR HA-17.  
 OS Clostridium botulinum.  
 OC Bacteria: Firmicutes; Bacillus/Clostridium group; Clostridiaceae;  
 OC Clostridium.  
 RN [1]  
 RN SEQUENCE FROM N.A.  
 RC MEDLINE: 96291660.  
 RA HENDERSON I., WHELAN S.M., DAVIS T.J., MINTON N.P.  
 RI "Genetic characterisation of the botulinum toxin complex of  
 Clostridium botulinum strain NGTC 2916.".  
 RL FEMS Microbiol. Lett. 140:151-158(1996).  
 RN [2]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN-CDC 328; (ATCC 43757);  
 RX MEDLINE: 98440323.  
 RA SANTOS-BUELGA J., COLLINS M.D., EAST A.K.  
 RI "Characterization of the genes encoding the botulinum neurotoxin  
 complex in a strain of Clostridium botulinum producing type B and F  
 neurotoxins.".  
 RL Curr. Microbiol. 37:312-318(1998).  
 DR EMBL: L42537; AAB42188.1; .  
 DR EMBL: Y13630; CAA73964.1; .  
 SQ SEQUENCE 146 AA: 17034 MW: C1990B1B CRC32:

Query Match 47.2% Score 59: DB 2: Length 146;  
 Best Local Similarity 58.3%; Pred. No. 3.84e+00;  
 Matches 7: Conservative 4: Mismatches 1: Indels 0: Gaps 0:

Db 133 DIYNSQWFKLE 144  
 QY 9 NIYNNNEAFKVE 20

RESULT 9  
 ID O83926 PRELIMINARY: PRT: 254 AA.

O83926;  
 AC 01-NOV-1998 (TRENBLrel. 08, Created)  
 DT 01-NOV-1998 (TRENBLrel. 08, Last sequence update)  
 DT 01-NOV-1998 (TRENBLrel. 10, Last annotation update)  
 DE FLAGELLAR BASAL-BODY ROD PRO. LIN (FLGG-1).  
 GN TPC960.  
 OS Treponema pallidum.  
 OC Bacteria; Spirochaetales; Spirochaetaceae; Treponema.  
 RN [1]  
 RN SEQUENCE FROM N.A.  
 RX MEDLINE: 98332770.  
 RA FRASER C.M., NORRIS S.J., WEINSTOCK G.M., WHITE O., SUTTON G.G.  
 RA DODSON R., GWINN M., HICKEY E.K., CLAYTON R., KEICHUM K.A.,  
 RA SODERREN E., HARDHAM J.M., MCLEOD M.P., SALZBERG S., PETERSON J.,  
 RA KHALAK H., RICHARDSON D., HOWELL J.K., CHIDAMBARAM M., UTTERBACK T.,  
 RA McDONALD L., ARTIACH P., BOWMAN C., COTTON M.D., FUJII C., GARLAND S.,  
 RA HATCH B., HORST K., ROBERTS K., WATTHEY L., WEIDMAN J., SMITH H.O.,  
 RA VENTER J.C.;  
 RI "Complete genome sequence of Treponema pallidum, the syphilis  
 spirochete.".  
 RL Science 281:375-388(1998).  
 RN [2]  
 RN SEQUENCE FROM N.A.  
 RA FRASER C.M., NORRIS S.J., WEINSTOCK G.M., WHITE O., SUTTON G.G.  
 RA DODSON R., GWINN M., HICKEY E.K., CLAYTON R., KEICHUM K.A.,  
 RA SODERREN E., HARDHAM J.M., MCLEOD M.P., SALZBERG S., PETERSON J.,  
 RA KHALAK H., RICHARDSON D., HOWELL J.K., CHIDAMBARAM M., UTTERBACK T.,  
 RA McDONALD L., ARTIACH P., BOWMAN C., COTTON M.D., FUJII C., GARLAND S.,  
 RA HATCH B., HORST K., ROBERTS K., WATTHEY L., WEIDMAN J., SMITH H.O.,  
 RA VENTER J.C.;  
 RI Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AE001264; AAC65615.1;  
 DR TIGR: TP0560;  
 DR PFAM: PF00460; flg\_bb\_rnd; .  
 KW Flagella.  
 SQ SEQUENCE 264 AA: 28597 MW: FC2A7F39 CRC32:  
 Query Match 47.2% Score 59: DB 2: Length 264;  
 Best Local Similarity 40.0%; Pred. No. 3.84e+00;  
 Matches 6: Conservative 4: Mismatches 5: Indels 0: Gaps 0:  
 Db 109 DGIYAYTRDGSFKVD 123  
 QY 6 EGTNIYNNNEAFKVE 20

RESULT 10  
 ID O19059 PRELIMINARY: PRT: 246 AA.  
 AC O19059  
 DT 01-NOV-1996 (TRENBLrel. 01, Created)  
 DT 01-NOV-1996 (TRENBLrel. 02, Last sequence update)  
 DT 01-NOV-1999 (TRENBLrel. 12, Last annotation update)  
 DE SIMILAR TO HYDRATASE-DEHYDROGENASE-EPIMERASE.  
 GN E04F6.3.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;  
 OC Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.  
 RN [1]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2.  
 RX MEDLINE: 94150718.  
 RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,  
 RA BONFIELD J., BURTON J.J., CONNELL M., COPSEY T., COOPER J., COULSON A.,  
 RA CRAXTON M., DEAR S., DU Z., DUBIN R., FAVELLO A., FULTON L.,  
 RA GARDNER A., GREEN P., KIRKINS J., HILLIER L., JIER M., JOHNSTON L.,  
 RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,  
 RA LIGHNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,  
 RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,  
 RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULTON J.,  
 RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,  
 RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;  
 RI "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
 elegans.".

RL Nature 368:32-38(1994).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN-BRISTOL N2:

RA PAULEY A.:

RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN-BRISTOL N2:

RA WATERSTON R.:

RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.

DR EMBL: U2843; AAA68358.1;

DR PFAM: PF01575; Maoc\_like.1;

SO SEQUENCE 298 AA; 32721 MW; F564B8FE CRC32;

Query Match

Best Local Similarity 47.28; Score 59; DB 5; Length 298;

Matches 7; Conservative 7; Mismatches 5; Indels 0; Gaps 0;

DB 32 ARAKEDLCYVYVNHEDFKV 50

QY 1 SSGKNEGINVYNNNEAFKV 19

RESULT 11

ID Q59078 PRELIMINARY:

AC Q59078; PRT: 290 AA.

DI 01-NOV-1996 (TRENBLrel. 01. Created)

DI 01-NOV-1996 (TRENBLrel. 01. Last sequence update)

DI 01-NOV-1996 (TRENBLrel. 01. Last annotation update)

DE BETA-AGARASE.1; PRECURSOR (EC 3.2.2.61).

US Alteromonas atlantica (Pseudomonas atlantica).

OC Bacteria; Proteobacteria; gamma subdivision; Alteromonadaceae;

PN Pseudalteromonas.

RN [1]

RP SEQUENCE FROM N.A.

RA CUAN C., RIGGS P.D., BENNER J.S., ZHU P., CHOU D., HALL I.S.,

RA KEATING C.P.:

RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.

CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF 1,4-BETA-D-GALACTOSIDIC LINKAGES

CC IN AGAROSE, GIVING THE TETRAMER AS THE PREDOMINANT PRODUCT.

DR EMBL: M73783; AAA91688.1;

DR SIGNAL: Hydrophobic; Glucosidase.

KT Signal: Hydrophobic; Glucosidase.

ET CHAIN 24 240 BETA-AGARASE.1

SO SEQUENCE 240 AA; 32795 MW; FAS03180 CRC32;

Query Match

Best Local Similarity 46.48; Score 58; DB 2; Length 290;

Matches 5; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

DB 187 KDAGSNVYNGEYR 201

QY 4 KNEGINVYNNNEAFKV 18

RESULT 12

ID Q9YW42 PRELIMINARY:

AC Q9YW42; PRT: 379 AA.

DI 01-MAY-1999 (TRENBLrel. 10. Created)

DI 01-MAY-1999 (TRENBLrel. 10. Last sequence update)

DI 01-MAY-1999 (TRENBLrel. 10. Last annotation update)

DE ORF MSV050 HYPOTHETICAL PROTEIN.

GN MSV050

OS Melanoplus sanguinipes entomopoxvirus.

OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Entomopoxvirinae;

PN Entomopoxvirus B.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-TUCSON:

RA MEDLINE: 99102612.

RA AFONSO C.L., TULMAN E.R., LU Z., OMA E., KUTISH G.F., ROCK D.L.:

RL "The genome of Melanoplus sanguinipes entomopoxvirus."

RL J. Virol. 73:533-552(1999).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN-TUCSON:

RA AFONSO C.L., TULMAN E.R., LU Z., OMA E., KUTISH G.F., ROCK D.L.:

RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.

DR EMBL: AF03866; AAC97519.1;

SO SEQUENCE 379 AA; 43956 MW; 72F41767 CRC32;

Query Match

Best Local Similarity 46.48; Score 58; DB 14; Length 379;

Matches 8; Conservative 8; Mismatches 2; Indels 0; Gaps 0;

DB 201 GTNGYNNMESFKI 213

QY 7 GINIYNNNEAFKV 19

RESULT 13

ID Q87980 PRELIMINARY:

AC Q87980; PRT: 432 AA.

DI 01-NOV-1998 (TRENBLrel. 08. Created)

DI 01-NOV-1998 (TRENBLrel. 08. Last sequence update)

DI 01-NOV-1998 (TRENBLrel. 08. Last annotation update)

DE HYPOTHETICAL 50.7 KD PROTEIN.

GN BBLPS1.07

OS Bordetella bronchiseptica.

OC Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;

PN Bordetella.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-CN7635E:

RA STEVENS K., CHURCHER C.M., BADOCK K.L.:

RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN-CN7635E:

RA PARHILL J., PRESTON A., MASKELL D.J., FARRELL B.G.:

RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.

DR EMBL: A207747; CAA07646.1;

KW Hypothetical protein.

SO SEQUENCE 432 AA; 50741 MW; 47891833 CRC32;

Query Match

Best Local Similarity 45.41; Score 56; DB 2; Length 412;

Matches 6; Conservative 6; Mismatches 4; Indels 0; Gaps 0;

DB 73 MBESSQIVYVGEDEF 87

QY 4 KNEGINVYNNNEAFKV 18

RESULT 14

ID O96205 PRELIMINARY:

AC O96205; PRT: 2206 AA.

DI 01-MAY-1999 (TRENBLrel. 10. Created)

DI 01-MAY-1999 (TRENBLrel. 10. Last sequence update)

DI 01-MAY-1999 (TRENBLrel. 10. Last annotation update)

DE HYPOTHETICAL 264.1 KD PROTEIN.

GN PFB0560K

OS Plasmodium falciparum.

OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE: 99021743.

RA GARDNER M.J., TETELIN H., CARUCCI D.J., CUMMINGS L.M., ARAVIND L.,

RA KOONIN E.V., SHALLOO S., MASON T., YU K., FUJII C., PEDERSON J.,

RA SHEN K., JING J., ASTON C., LAI Z., SCHWARTZ D.C., PERTEA M.,

RA SALZBERG S., ZHOU L., SUTTON G.G., CLAYTON R., WHITE C., SMITH H.G.,

RA FRASER C.M., ADAMS M.D., VENTER J.C., HOFFMAN S.L.:

RL Chromosome 2 sequence of the human malaria parasite Plasmodium

RL falciparum.

RL Science 282:1126-1132(1998).

```

DP EMBL: AEO01403; AAC71901.1; -
KW Hypothetical protein.
SQ SEQUENCE 2206 AA: 264100 MW: AEE13480 CRC32:

Query Match 45.6% Score 57: DB 5: Length 2206;
Best Local Similarity 35.7% Pred. No. 9 07e+00;
Matches 5: Conservative 6: Mismatches 3: Indels 0: Gaps 0:

DE 363 NNENIDVYNDKDF 376
QY 4 KNEGNIYNNNEAF 17

RESULT 15
ID 077336 PRELIMINARY: PRI: 4550 AA.
AC 077336:
DI 01-NOV-1998 (TRENBLrel. 08, Created)
DI 01-NOV-1998 (TRENBLrel. 08, Last sequence update)
DI 01-NOV-1999 (TRENBLrel. 12, Last annotation update)
DE PFC0425W PROTEIN.
GN PFC0425W.
OS Plasmodium falciparum.
OC Eukaryota: Alveolata: Apicomplexa: Haemosporida: Plasmodium.
RN [1]
RP SEQUENCE FROM N.A.
RA MUNGALL K., LAWSON D., BARRELL B.:
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: 798547; CAB11211.1; -
SQ SEQUENCE 4550 AA: 534109 MW: CC698096 CRC32:

Query Match 45.6% Score 57: DB 5: Length 4550;
Best Local Similarity 18.8% Pred. No. 9 07e+00;
Matches 3: Conservative 9: Mismatches 4: Indels 0: Gaps 0:

DE 831 EDDDDIYDDDDNFED 846
QY 5 NEGNIYNNNEAFKE 20

```

Search completed: Mon Jun 19 16:15:11 2000  
 Job time : 12 secs.

\*\*\*\*\*  
M E R E F (TM)  
\*\*\*\*\*  
Release 3.1A John F. Collins, BioComputing Research Unit,  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd  
MPArch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Mon Jun 19 16:17:34 2000 MasPar time 3.82 Seconds  
Tabular output not generated.  
80,507 Million cell updates/sec

Title: >US-09-142-524A-13  
Description: (1-13) from US09142524A.pap  
Perfect Score: 94  
Sequence: VFIRKVSNIIVHG 13  
Scoring table: PAM 150  
Gap 15  
Searched: 189663 seqs, 23686106 residues  
Post-processing: Minimum Match 0%  
Listing first 45 summaries  
Database: a-geneseq35  
:igoneseqp  
Statistics: Mean 18.336: Variance 56.950: scale 0.322

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query | Water Length | IC | Description                    | Pred. No. |
|------------|-------|-------|--------------|----|--------------------------------|-----------|
| 1          | 94    | 100.0 | 45           | 1  | R89291 Japanese cedar pollen   | 4.88e-03  |
| 2          | 94    | 100.0 | 20           | 1  | R45552 Cry j 1 pollen allergen | 2.88e-03  |
| 3          | 94    | 100.0 | 20           | 1  | R82501 Cry j 1 Japanese Cedar  | 2.88e-03  |
| 4          | 94    | 100.0 | 30           | 1  | W44583 T-cell epitope peptide  | 2.88e-03  |
| 5          | 94    | 100.0 | 80           | 1  | W27369 Multi-epitope peptide   | 2.88e-03  |
| 6          | 94    | 100.0 | 195          | 1  | W27370 Multi-epitope peptide   | 2.88e-03  |
| 7          | 94    | 100.0 | 134          | 1  | W27371 Multi-epitope peptide   | 2.88e-03  |
| 8          | 94    | 100.0 | 353          | 1  | R75388 Japanese cedar pollen   | 2.88e-03  |
| 9          | 94    | 100.0 | 353          | 1  | R81587 Cedar pollen allergen   | 2.88e-03  |
| 10         | 94    | 100.0 | 374          | 1  | R31937 Cry j 1                 | 2.88e-03  |
| 11         | 94    | 100.0 | 374          | 1  | R82490 Cry j 1 Japanese Cedar  | 2.88e-03  |
| 12         | 94    | 100.0 | 374          | 1  | R45541 Cry j 1 pollen allergen | 2.88e-03  |
| 13         | 94    | 100.0 | 374          | 1  | R60166 Japanese cedar pollen   | 2.88e-03  |
| 14         | 82    | 87.2  | 15           | 1  | W57750 Residues 106-120 of Cr  | 7.22e-02  |
| 15         | 69    | 73.4  | 20           | 1  | R45553 Cry j 1 pollen allergen | 2.12e-02  |
| 16         | 69    | 73.4  | 20           | 1  | R82502 Cry j 1 Japanese Cedar  | 2.12e-02  |
| 17         | 67    | 71.3  | 370          | 1  | R45578 Jun v 1                 | 3.57e-02  |
| 18         | 64    | 68.1  | 9            | 1  | W57751 Residues 109-117 of Cr  | 7.46e-02  |
| 19         | 64    | 68.1  | 31           | 1  | W27372 Multi-epitope peptide   | 7.46e-02  |
| 20         | 64    | 68.1  | 31           | 1  | W27373 Multi-epitope peptide   | 7.46e-02  |
| 21         | 61    | 64.9  | 20           | 1  | W42373 T-cell epitope peptide  | 1.57e-01  |
| 22         | 61    | 64.9  | 354          | 1  | W04344 Chamaecyparis obtusa p  | 1.57e-01  |
| 23         | 61    | 64.9  | 354          | 1  | W42341 Japanese cypress polye  | 1.57e-01  |

|    |    |      |      |   |                               |          |
|----|----|------|------|---|-------------------------------|----------|
| 24 | 61 | 64.9 | 357  | 1 | R45577 Jun s 1                | 1.57e-01 |
| 25 | 61 | 64.9 | 375  | 1 | W04345 Chamaecyparis obtusa p | 1.57e-01 |
| 26 | 56 | 59.6 | 292  | 1 | R04895 Penicillinae-Insulin   | 5.30e-01 |
| 27 | 56 | 59.6 | 408  | 1 | W94245 A. orientalis glycosyl | 5.30e-01 |
| 28 | 56 | 59.6 | 408  | 1 | W94245 A. orientalis glycosyl | 5.30e-01 |
| 29 | 54 | 57.4 | 367  | 1 | W56279 Babesia microti BMN1-3 | 8.54e-01 |
| 30 | 54 | 57.4 | 367  | 1 | W56301 Babesia microti antige | 8.54e-01 |
| 31 | 52 | 55.3 | 192  | 1 | R89533 Hepatitis C virus iscl | 1.37e-02 |
| 32 | 52 | 55.3 | 192  | 1 | R89662 Hepatitis C virus enve | 1.37e-02 |
| 33 | 51 | 54.3 | 40   | 1 | R70757 Spider venom FIL-376 N | 1.73e-02 |
| 34 | 51 | 54.3 | 308  | 1 | W32430 Mycobacterium tubercul | 1.73e-02 |
| 35 | 51 | 54.3 | 309  | 1 | W32362 Mycobacterium tubercul | 1.73e-02 |
| 36 | 51 | 54.3 | 309  | 1 | W61665 Mycobacterium tubercul | 1.73e-02 |
| 37 | 51 | 54.3 | 309  | 1 | W61665 M. tuberculosis immuno | 1.73e-02 |
| 38 | 51 | 54.3 | 857  | 1 | W77702 Staphylococcus aureus  | 1.73e-02 |
| 39 | 50 | 53.2 | 316  | 1 | W77356 PSB (WBPf) protein in  | 2.18e-02 |
| 40 | 50 | 53.2 | 552  | 1 | W16135 Omega-cyclohexane fatt | 2.18e-02 |
| 41 | 50 | 53.2 | 672  | 1 | W96657 Staphylococcus aureus  | 2.18e-02 |
| 42 | 50 | 53.2 | 866  | 1 | W96657 Staphylococcus aureus  | 2.18e-02 |
| 43 | 50 | 53.2 | 908  | 1 | R33429 E. coli heat shock pro | 2.18e-02 |
| 44 | 49 | 52.1 | 1565 | 1 | W18305 Photorhabdus luminesce | 2.74e-02 |
| 45 | 49 | 52.1 | 1565 | 1 | W56568 Toxin TccB, encoded by | 2.74e-02 |

ALIGNMENTS

RESULT 1 standard; peptide: 15 AA.  
ID R89291 standard; peptide: 15 AA.  
AC R89291 standard; peptide: 15 AA.  
DE 12-MAR-1995 (first entry)  
DI Japanese cedar pollen allergen Cryj 1 derived T-cell epitope peptide.  
DF Japanese cedar; pollen allergen; Cryj 1; T-cell epitope; peptides.  
KW prevention; treatment: cryptomeria pollenosis.  
OS Cryptomeria japonica.  
PN J07118295-A.  
PD 09-MAY-1995.  
PF 20-OCT-1993; 262626.  
PR 20-OCT-1993; JP-262626.  
PA (MEIP) MEIJI MILK PROD CO LI.  
DR WPI: 95-203834/27.  
PI New cryptomeria pollen allergen T-cell epitope peptide - used for:  
PI prevention, treatment and investigation of Japanese cedar pollenosis  
PS Claim 5; page 2; 8pp: Japanese.  
CC R33388 is the Japanese cedar pollen allergen Cryj 1, from which the  
CC T-cell epitope peptides R89289-R89295 were derived. The peptides  
CC can be used for the prevention and treatment of cryptomeria  
CC pollenosis, and also for the investigation of pollenosis.  
SQ Sequence 15 AA.

Query Match: 100.0% Score 94: DB 1: Length 15.  
Best Local Similarity 100.0% Pred. No: 2.88e-03.  
Matches 13: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

DB 3 VFIRKVSNIIVHG 15  
| | | | | | | | | | | | | | | |  
QY 1 VFIRKVSNIIVHG 13

RESULT 2 standard; Protein: 20 AA.  
ID R45552 standard; Protein: 20 AA.  
AC R45552 standard; Protein: 20 AA.  
DI 13-JUL-1994 (first entry)  
DE Cry j 1 pollen allergen peptide CJI-11.  
KW Japanese cedar; detection: a.ergy; treatment: diagnosis:  
KW T cell epitope; sensitivity.  
OS Cryptomeria japonica.  
PN W0401560-A.  
PD 20-JAN-1994.  
PF 15-JAN-1993; U00139.  
PR 10-JUL-1992; WC-U05661.  
PR 01-SEP-1992; US-938990.  
PA (IMMU) IMMUNOLOGIC PHARM CORP.  
PI Bond JF, Garraan RD, Griffith JJ, Kuo M, Pollock J:

DR WPI: 94-03566/24.  
 PR Antigen derived from Japanese cedar pollen allergen Cry j I -  
 PA contain at least two T cell epitope(s), used to treat or diagnose  
 PT allergy.  
 PS Claim 1: Fig 13: 137pp; English.  
 CC The sequence is that of an isolated peptide of the Japanese cedar  
 pollen allergen Cry j I (amino acids 101-120). The peptide, CJI-11,  
 can be used for the treatment and diagnosis of allergies associated  
 with Japanese cedar pollen. It has enhanced therapeutic properties  
 but reduced side effects compared to naturally occurring allergens.  
 SQ Sequence 20 AA:

Query Match 100.0%; Score 94; DB 1; Length 20;

Best Local Similarity 100.0%; Pred No. 2,89e-03;  
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 8 VFIRKRVSNVIING 20  
 QY 1 VFIRKRVSNVIING 13

RESULT 3  
 ID R82501 standard; Protein: 20 AA.

AC R82501;  
 DE Cry j I Japanese Cedar pollen allergen peptide fragment (CJI-11).  
 KW Cry j I: Japanese cedar pollen allergen; modified: drug production;  
 KW Allergy; Cryptomeria japonica.  
 CS Cryptomeria japonica.  
 PN W09527786-A1.  
 FN 19-OCT-1995.  
 PR 06-APR-1995; U04249.  
 PR 08-APR-1994; US-226248.  
 PR 06-DEC-1994; US-150225.  
 PA (IMMUNO) IMMUNOLOGIC PHARM CORP.  
 PI Chen X., Evans S., Fahren HM., Kuo X., Powers SP.  
 PT Shaked Z.  
 PR WPI: 95-36530/147.  
 PT Modified Cryptomeria japonica (Cry j) I peptide(s) - useful for  
 treating allergy to Japanese cedar pollen allergen or  
 immunologically cross reactive allergens.  
 PS Disclosure: Figure 2: 60pp; English.  
 CC Novel peptides of Cry j I have been modified as a part of a  
 preformulation scheme to develop an optimised drug product for  
 therapeutic treatment of humans suffering from allergy to Japanese  
 cedar pollen allergen or an allergen which is immunologically cross  
 reactive with Japanese cedar pollen allergen. Such modified peptides  
 possess certain characteristics which render them particularly  
 suitable for drug product formulation. Peptide fragments of Cry j I  
 modified and recombined, are given in R8249-182625. This peptide  
 fragment corresponds to amino acids 101-120 of the allergen mature  
 protein.

SQ Sequence 20 AA:

Query Match 100.0%; Score 94; DB 1; Length 20;

Best Local Similarity 100.0%; Pred No. 2,89e-03;  
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 6 VFIRKRVSNVIING 20  
 QY 1 VFIRKRVSNVIING 13

RESULT 4  
 ID W44683 standard; peptide: 20 AA.

AC W44683;  
 DE 01-MAY-1998 (first entry)  
 DE T-cell epitope peptide #2 of sugi pollen antigen.  
 KW T-cell epitope, sugi pollen antigen; sugi pollinosis.  
 CS Synthetic.  
 PN J1000780-A  
 FN 13-JAN-1998

PF 24-JUN-1996: 163287.  
 PR 24-JUN-1996: JP-163287.  
 PA (DAIL) DAICEL CHEM IND LTD.  
 PA (MEIJ) MEIJI SEIKA KAISHA LTD.  
 DR WPI: 98-133630/13.

PT T-cell epitope peptide of sugi pollen antigen - useful in the  
 treatment of sugi pollinosis.  
 PS Claim 1: Page 4: 14pp; Japanese.  
 CC T-cell epitope peptides W44682-83 and their derivatives react with  
 sugi pollinosis patient peripheral blood T lymphocytes. A composition  
 prepared by combining at least 2 of the above peptides and/or their  
 CC derivatives is used for the prevention and treatment of sugi pollinosis.  
 SQ Sequence 30 AA:

Query Match 100.0%; Score 94; DB 1; Length 30;

Best Local Similarity 100.0%; Pred No. 2,89e-03;  
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 8 VFIRKRVSNVIING 20  
 QY 1 VFIRKRVSNVIING 13

RESULT 5

ID W27369 standard; peptide: 80 AA.

AC W27369;  
 DE 24-MAR-1998 (first entry)  
 DE Multi-epitope peptide used as immunotherapeutic agent #1.  
 KW Multi-epitope peptide; immunotherapeutic agent; allergic disease;  
 KW T-cell epitope region; allergen; lymphocyte; immunoglobulin E.  
 OS Synthetic.  
 PN W09732600-A1.  
 PD 12-SEP-1997.  
 PR 10-MAR-1997: J00740.  
 PR 10-MAR-1996: JP-080702.  
 PA (MEIP) MEIJI MILK PROD CO LTD.  
 PI Cairiki K., Iwara A., Kino K., Kume A., Sone T.  
 PR WPI: 97-470495/43.  
 PT Peptide is immunotherapeutic agent to treat allergic diseases.  
 PT contains multi-epitope peptide containing T cell epitope region  
 from different allergens.  
 PS Claim 6: Page 31: 58pp; Japan.  
 CC The present sequence represents a multi-epitope peptide which is used as  
 a new immunotherapeutic agent. It comprises T cell epitope regions from 2  
 or more different allergens (preferably linked via arginine or lysine  
 dimers), where the T cell epitope regions have a positivity index  
 greater than 100 as measured in a patient group responding to the  
 CC allergen; have at least 70% reactivity with lymphocytes from patients  
 CC responding to the allergen; and are not reactive with immunoglobulin E.  
 CC (IgE) antibodies from patients responsive to the allergen. The agent can  
 be used to prevent and treat a wide variety of allergic diseases, such as  
 CC desensitisation. Side effects and those mediated by IgE are reduced.  
 SQ Sequence 80 AA:

Query Match 100.0%; Score 94; DB 1; Length 80;

Best Local Similarity 100.0%; Pred No. 2,89e-03;  
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 16 VFIRKRVSNVIING 28  
 QY 1 VFIRKRVSNVIING 13

RESULT 6

ID W27370 standard; peptide: 105 AA.

AC W27370;  
 DE 24-MAR-1998 (first entry)  
 DE Multi-epitope peptide used as immunotherapeutic agent #2.  
 KW Multi-epitope peptide; immunotherapeutic agent; allergic disease;  
 KW T-cell epitope region; allergen; lymphocyte; immunoglobulin E.  
 OS Synthetic.  
 PN W09732600-A1.  
 PD 12-SEP-1997.

```

PF 10-MAR-1997: J00740.
PR 10-MAR-1996: JP-080702.
PA (MEIP ) MEIJI MILK PROD CO LTD.
PI Dairiki K, Iwama A, Kino K, Kume A, Sone T;
PW: 97-470495/43.
BI Peptide immunotherapeutic agent to treat allergic diseases -
PI contains multi-epitope peptide containing T cell epitope regions
from different allergens
PS Claim 6: Page 31: 58pp; Japanese.
CC The present sequence represents a multi-epitope peptide which is used as
CC a new immunotherapeutic agent. It comprises T cell epitope regions from 2
CC or more different allergens (preferably linked via arginine or lysine
CC dimers), where the T cell epitope regions have a positivity index
CC greater than 100 as measured in a patient group responding to the
CC allergen; have at least 70% reactivity with lymphocytes from patients
CC responding to the allergen; and are not reactive with immunoglobulin E
CC (IgE) antibodies from patients responsive to the allergen. The agent can
CC be used to prevent and treat a wide variety of allergic diseases, e.g. by
CC desensitisation. Side effects, e.g. those mediated by IgE, are reduced.
SV Sequence 105 AA:

Query Match 100.0%; Score 94; DB 1; Length 105;
Best Local Similarity 100.0%; Pred. No. 2.88e-03;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 16 VFIKRVSNVIHG 28
QY 1 VFIKRVSNVIHG 13

RESULT 7
ID W27371 standard; peptide; 134 AA.
AC W27371 (first entry)
DE Multi-epitope peptide used as immunotherapeutic agent #3.
KW Multi-epitope peptide; immunotherapeutic agent; allergic disease;
KW T-cell epitope region; allergen; lymphocyte; immunoglobulin E.
CS Synthetic.
PN WC9732600-A1.
PC 12-SEP-1997.
PF 10-MAR-1997: J00740.
PR 10-MAR-1996: JP-080702.
PA (MEIP ) MEIJI MILK PROD CO LTD.
PI Dairiki K, Iwama A, Kino K, Kume A, Sone T;
PW: 97-470495/43.
BI Peptide immunotherapeutic agent to treat allergic diseases -
PI contains multi-epitope peptide containing T cell epitope regions
from different allergens
PS Claim 6: Page 32: 58pp; Japanese.
CC The present sequence represents a multi-epitope peptide which is used as
CC a new immunotherapeutic agent. It comprises T cell epitope regions from 2
CC or more different allergens (preferably linked via arginine or lysine
CC dimers), where the T cell epitope regions have a positivity index
CC greater than 100 as measured in a patient group responding to the
CC allergen; have at least 70% reactivity with lymphocytes from patients
CC responding to the allergen; and are not reactive with immunoglobulin E
CC (IgE) antibodies from patients responsive to the allergen. The agent can
CC be used to prevent and treat a wide variety of allergic diseases, e.g. by
CC desensitisation. Side effects, e.g. those mediated by IgE, are reduced.
SV Sequence 134 AA:

Query Match 100.0%; Score 94; DB 1; Length 134;
Best Local Similarity 100.0%; Pred. No. 2.88e-03;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 16 VFIKRVSNVIHG 28
QY 1 VFIKRVSNVIHG 13

RESULT 8
ID R75368 standard; protein; 353 AA.
AC R75368

```

```

DT 12-MAR-1996 (first entry)
DE Japanese cedar pollen allergen Cryj I.
KW Japanese cedar; pollen allergen; Cryj I; T-cell epitope; peptides;
KW prevention; treatment; cryptomeria pollinosis.
OS Cryptomeria japonica.
FH Key Location/Qualifiers
FT peptide 51..175 /note= "T-cell epitope peptide"
FT peptide 91..105 /note= "T-cell epitope peptide"
FT peptide 106..120 /note= "T-cell epitope peptide"
FT peptide 146..160 /note= "T-cell epitope peptide"
FT peptide 211..225 /note= "T-cell epitope peptide"
FT peptide 326..340 /note= "T-cell epitope peptide"
FT peptide 335..346 /note= "T-cell epitope peptide"
PN J07118295-A.
PD 09-MAY-1995.
PF 20-OCT-1993: 262626.
PR 20-OCT-1993: JP-262626.
PA (MEIP ) MEIJI MILK PROD CO LTD.
DR WPI: 95-203834/27.
PI New cryptomeria pollen allergen T-cell epitope peptide - used for
PI prevention, treatment and investigation of Japanese cedar pollinosis.
PS Disclosure: Figs 1-2: 8pp; Japanese.
CC R75388 is the Japanese cedar pollen allergen Cryj I, from which the
CC T-cell epitope peptides R892, R89295 were derived. The peptides
CC can be used for the prevention and treatment of cryptomeria
CC pollinosis, and also for the investigation of pollinosis.
SQ Sequence 353 AA:

Query Match 100.0%; Score 94; DB 1; Length 353;
Best Local Similarity 100.0%; Pred. No. 2.88e-03;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 108 VFIKRVSNVIHG 120
QY 1 VFIKRVSNVIHG 13

RESULT 9
ID R81587 standard; Protein; 353 AA.
AC R81587.
DE 24-MAY-1996 (first entry)
DE Cedar pollen allergen B.
KW Cedar; pollen allergen; immunoglobulin E; IgE; T-cell epitope;
KW antibody; pollinosis; therapy; immunotherapy.
OS Cryptomeria japonica.
PN EP-700929-A2.
PC 13-MAR-1996.
PF 08-SEP-1995: 306295.
PR 10-SEP-1994: JP-242137.
PR 14-JUL-1995: JP-200221.
PR 14-JUL-1995: JP-200204.
PA (HAYB ) HAYASHIBARA SEIBUTSU KAGAKU.
PI Hino K, Saito S, Taniguchi Y;
DR WPI: 96-140976/15.
PT New peptide(s) derived from cedar pollen allergens - activate
PT allergen-specific T-cells, but not allergen-specific IgE antibodies.
PI used for treating cedar pollinosis
PS Claim 5: Page 31-32: 36pp; English.
CC Synthetic peptides based on portions of cedar pollen allergens A
CC (R81586) and B (R81587) were tested for their ability to activate
CC cedar allergen-specific T-cells, but not allergen-specific IgE
CC antibodies. 6 peptides (R81580-R81585) were identified as T-cell
CC epitopes. These peptides, plus subsequences (R81573-79) essential
CC for T-cell recognition, and homologous peptides (R81588-96) can
CC be used as immunotherapeutic agents to treat or prevent cedar
CC pollinosis, avoiding side-effects such as anaphylaxis.

```

SQ Sequence 353 AA:  
 Query Match 100.0% Score 94: DB 1: Length 353:  
 Best Local Similarity 100.0% Pred. No. 2,88e-03:  
 Matches 13: Conservative 0: Mismatches 0: Indels 0: Gaps 0:  
 DE 108 VFIKRVSNV:HG 120  
 IIII IIIII  
 QY 1 VFIKRVSNV:HG 13  
 IIII IIIII  
 RESULT 10  
 ID R31937 standard: Protein: 374 AA.  
 AC R31937:  
 DT 03-JUN-1993 (first entry)  
 DE Cry j 1: Japanese cedar pollen; allergen: antigen; allergy: B cell; T cell.  
 KW Cryptomeria japonica.  
 FH Key Location/Qualifiers  
 FT Peptide 1..21 /note= "signal peptide"  
 FT protein 22..374 /note= "mature Cry j 1"  
 FT WO9301213-A.  
 DT 21-JAN-1993:  
 DE 10-JUL-1992: U05661.  
 DE 12-JUL-1991: US-729134.  
 DE 15-JUL-1991: US-730452.  
 DA (IMMUNO) IMMUNOLOGIC PHARM CORP.  
 PL Bond JF, Griffith LC, Pollock J.  
 DR WP1: 93-045434/05.  
 DR N-PSDB: Q35304.  
 PI Nucleic acid sequence encoding Cryptomeria japonica allergen -  
 PT for the diagnosis treatment and prevention of allergic reactions  
 PT to Japanese cedar pollen  
 PS Claim 1: Page 42: 69pp: English.  
 CC Fresh pollen and staminate cone samples were collected from a single  
 CC Cryptomeria japonica (Japanese cedar) tree. RNA was prep'd. and used  
 CC to synthesize cDNA. The cDNA was subjected to successive rounds of  
 CC PCR to yield a full length Cry j 1 clone. Cry j 1 or an antigenic  
 CC fragment of it may be used for detecting, treating and preventing an  
 CC allergic response to Japanese cedar pollen allergen. It is capable of  
 CC modifying both the B and T cell response to Cry j 1 and T cell response  
 CC to a Cry j 1 antigen.  
 SQ Sequence 374 AA:  
 Query Match 100.0% Score 94: DB 1: Length 374:  
 Best Local Similarity 100.0% Pred. No. 2,88e-03:  
 Matches 13: Conservative 0: Mismatches 0: Indels 0: Gaps 0:  
 DE 129 VFIKRVSNV:HG 141  
 IIII IIIII  
 QY 1 VFIKRVSNV:HG 13  
 IIII IIIII  
 RESULT 11  
 ID R82490 standard: Protein: 374 AA.  
 AC R82490:  
 DT 15-APR-1996 (first entry)  
 DE Cry j 1 Japanese Cedar Pollen allergen.  
 KW Cry j 1: Japanese cedar pollen allergen; modified: drug production;  
 KW allergy: Cryptomeria japonica.  
 FH Key Location/Qualifiers  
 FT Peptide 1..21 /note= "signal peptide"  
 FT protein 22..374 /note= "mature Cry j 1"  
 FT WO9527786-A.  
 DT 19-OCT-1995:  
 DE 06-APR-1995: U04249.  
 DE 08-APR-1994: US-226248.  
 DE 06-DEC-1994: US-350425.  
 DA (IMMUNO) IMMUNOLOGIC PHARM CORP.  
 PL Chen X, Evans S, Franzen HM, Kuo M, Powers SP,  
 PI Skaked Z.  
 DR WP1: 95-366391/47.  
 DR N-PSDB: T04249.

PT Modified Cryptomeria japonica (Cry j 1) peptide(s) - useful for  
 PT treating allergy to Japanese cedar pollen allergen or  
 PT immunologically cross reactive allergens  
 PS Disclosure: Figure 1: 60pp: English.  
 CC Novel peptides of cry j 1 have been modified as a part of a  
 CC preformulation scheme to develop an optimised drug product for  
 CC therapeutic treatment of human suffering from allergy to Japanese  
 CC cedar pollen allergen or an allergen which is immunologically cross  
 CC reactive with Japanese cedar pollen allergen. Such modified peptides  
 CC possess certain characteristics which render them particularly  
 CC suitable for drug product formulation. Peptide fragments of Cry j 1:  
 CC modified and unmodified, are given in R82491-R82525.  
 SQ Sequence 374 AA:  
 Query Match 100.0% Score 94: DB 1: Length 374:  
 Best Local Similarity 100.0% Pred. No. 2,88e-03:  
 Matches 13: Conservative 0: Mismatches 0: Indels 0: Gaps 0:  
 DE 129 VFIKRVSNV:HG 141  
 IIII IIIII  
 QY 1 VFIKRVSNV:HG 13  
 IIII IIIII  
 RESULT 12  
 ID R45541 standard: Protein: 374 AA.  
 AC R45541:  
 DT 13-JUL-1994 (first entry)  
 DE Cry j 1 pollen allergen  
 DE Japanese cedar; detection: allergy; treatment: diagnosis;  
 KW cell epitope; sensitivity.  
 CS Cryptomeria japonica.  
 FH Key Location/Qualifiers  
 FT Peptide 1..21 /note= "signal peptide"  
 FT peptide 22..374 /note= "mature peptide"  
 FT WO9401560-A.  
 DT 20-JAN-1994:  
 DE 15-JAN-1993: U00139.  
 DE 10-JUL-1992: WO-U05661.  
 DE 01-SEP-1992: US-938990.  
 DA (IMMUNO) IMMUNOLOGIC PHARM CORP.  
 PL Bond JF, Garman RD, Griffith LC, Pollock J.  
 DR WP1: 94-035066/04.  
 DR N-PSDB: Q55271.  
 PI Antigens derived from Japanese cedar pollen allergen Cry j 1:  
 PT contain at least two T cell epitope(s). used to treat or diagnose  
 PT allergy  
 PS Disclosure: Fig 4: 137pp: English  
 CC The sequence is that of the Japanese cedar pollen allergen  
 CC Cry j 1 which contains at least two T cell epitopes. Peptide  
 CC antigens derived from it can be used for the treatment and  
 CC diagnosis of allergies associated with Japanese cedar pollen.  
 CC The peptides have enhanced therapeutic properties but reduced  
 CC side effects compared to naturally occurring allergens.  
 SQ Sequence 374 AA:  
 Query Match 100.0% Score 94: DB 1: Length 374:  
 Best Local Similarity 100.0% Pred. No. 2,88e-03:  
 Matches 13: Conservative 0: Mismatches 0: Indels 0: Gaps 0:  
 DE 129 VFIKRVSNV:HG 141  
 IIII IIIII  
 QY 1 VFIKRVSNV:HG 13  
 IIII IIIII  
 RESULT 13  
 ID R50166 standard: Protein: 374 AA.  
 AC R50166:  
 DT 24-MAR-1995 (first entry)  
 DE Japanese cedar pollen antigen Cryj1  
 KW Japanese cedar; pollen antigen; allergen: Cryj1; sugi; pollinosis.  
 CS Cryptomeria japonica.

```

DE Key Location/Qualifiers
DE peptide /label: signal_peptide
DE 22: 374
DE prole.n /label: mature_CryJ:
DE J061977748.A.
DE 19-JUL-1994.
DE 07-JAN-1993: C01115.
DE 07-JAN-1993: C01116.
DE (METS: METS: SEIKA KAISHA.
DE WP1: 94-26866/13
DE N-ESCU: 072601.
DE Sugi (Japanese cedar) pollen antigen CryJ1 is useful for
DE diagnosis, treatment and prevention of sugi pollinosis.
DE Claim 2: Page 57: 9pp: Japanese.
DE The coding sequence for the Japanese cedar ("sugi") pollen allergen
DE CryJ1 was isolated from a cDNA library prepared from polyA mRNA. All
DE or part of the CryJ1 protein can be used for diagnosis, treatment
DE and prevention of sugi pollinosis.
DE Sequence 15 AA:
Query Match 100.0% Score 94: DB 1: Length 374:
Best Local Similarity 100.0% Pred. No. 2: 896-03:
Matches 13: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

```

```

DE 129 VKKRVSNVTHG 14:
DE 111: 111
DE 1 VKKRVSNVTHG 13
DE 1 VKKRVSNVTHG 13

```

```

RESULT 14
DE W57750 standard: peptide: 15 AA:
DE W57750
DE 17-SEP-1998 (first entry)
DE Residues 165-220 of Cry J1.
DE Cry J1: Japanese Cedar pollen antigen; allergy: immunotherapy;
DE HLA class II molecule.
DE Cryptomeria japonica.
DE W098209.2.A.
DE 22-MAY-1998.
DE 12-N-V-1997: C04129.
DE (METS: METS: HLK PRD C3 LTD.
DE Gairiki K, Kuno K, Kume A, Soga T:
DE WP1: 97-27517/26.
DE Peptides derived from Japanese cedar pollen antigens are
DE immunotherapeutic agents useful for allergy treatment and typing
DE HLA class II molecules in allergy sufferers
DE Claim 1: Page 26: 5pp: Japanese.
DE This sequence represents residues 165-220 of the Cry J1 protein, and
DE is a peptide of the invention. The peptides are derived from Japanese
DE cedar pollen antigens and are used as immunotherapeutic agents in the
DE treatment of allergy. The peptides can be used for identification and
DE typing of the particular HLA class II molecules in an allergy sufferer,
DE and also for peptide immunotherapy of an allergy. Using these peptides
DE the immunotherapy can be targeted more specifically to the requirements
DE of the individual patient, allowing more effective treatment of an
DE allergy, including those patients for whom treatment with a conventional
DE immunotherapeutic agent is ineffective.
DE Sequence 15 AA:

```

```

Query Match 87.2% Score 82: DB 1: Length 15:
Best Local Similarity 92.3% Pred. No. 7: 226-02:
Matches 12: Conservative 0: Mismatches 1: Indels 0: Gaps 0:

```

```

DE 3 VKKRVSNVTHG 15
DE 111: 111
DE 1 VKKRVSNVTHG 13

```

```

RESULT 15
DE R45551 standard: Protein: 20 AA:
DE R45551

```

```

DE 13-JUL-1994 (first entry)
DE Cry J1 pollen allergen peptide C01-12.
DE Japanese cedar: detection; allergy: treatment; diagnosis;
DE I cell epitope; sensitivity.
DE Cryptomeria japonica.
DE W09401560.A.
DE 20-JAN-1994.
DE 15-JAN-1993: C00139.
DE 10-JUL-1992: W0305661.
DE 01-SEP-1992: US-918990.
DE (METS: IMMUNOLOGIC PHARM CORP.
DE Bond JF, Garman RD, Griffith M, Kuo M, Pollock J:
DE WP1: 94-035086/04.
DE Antigens derived from Japanese cedar pollen allergen Cry J1
DE contain at least two T cell epitope(s) used to treat of Japanese
DE allergy.
DE Claim 1: Fig 13: 137pp: English.
DE The sequence is that of an isolated peptide of the Japanese cedar
DE pollen allergen Cry J1 (amino acids 111-133). The peptide, which
DE can be used for the treatment and diagnosis of allergies associated
DE with Japanese cedar pollen, it has enhanced therapeutic properties
DE but reduced side effects for use to naturally occurring antigens.
DE Sequence 20 AA:

```

```

Query Match 73.4% Score 59: DB 1: Length 20:
Best Local Similarity 100.0% Pred. No. 2: 126-01:
Matches 10: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

```

```

DE 1 KRVSNVTHG 10
DE 111: 111
DE 4 KRVSNVTHG 13

```

```

Search completed: Mon Jun 19 11:11:40 2000
Job time : 6 secs.

```



.....  
[W][O][R][D] [I][M]  
.....

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Tue Jun 20 13:37:24 2000: MasPar time 2.28 Seconds  
Tabular output not generated. 82.449 Million cell updates/sec

Title: >US-09-142-524A-13  
Description: (1-13) from US09142524A.pap  
Perfect Score: 94  
Sequence: 1 VFIKRVSNV:HG 13

Scoring table: PAM 150  
Gap 15

Searched: 145341 seqs, 14437480 residues

Post-processing: Minimum Match 0%  
Listing first: 45 summaries

Database: a-issued  
1:5A\_COMB 2:5B\_COMB 3:6\_COMB 4:PCI\_COMB 5:backfiles1

Statistics: Mean 17.285; Variance 56.805; scale 0.304

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description                        | Pred. No. |
|------------|-------|-------------|--------|-------|------------------------------------|-----------|
| 1          | 56    | 59.6        | 235    | 1     | US-08-459- Sequence 4, Applicatio  | 3.56e-01  |
| 2          | 56    | 59.6        | 235    | 2     | US-08-459- Sequence 4, Applicatio  | 3.56e-01  |
| 3          | 56    | 59.6        | 408    | 2     | US-09-120- Sequence 2, Applicatio  | 3.56e-01  |
| 4          | 56    | 59.6        | 408    | 2     | US-08-926- Sequence 2, Applicatio  | 3.56e-01  |
| 5          | 52    | 55.3        | 132    | 4     | PCT-US95-1 Sequence 80, Applicati  | 9.20e-01  |
| 6          | 52    | 55.3        | 132    | 1     | US-08-086- Sequence 80, Applicati  | 9.20e-01  |
| 7          | 52    | 55.3        | 132    | 2     | US-08-930- Sequence 80, Applicati  | 9.20e-01  |
| 8          | 52    | 55.3        | 132    | 2     | US-08-458- Sequence 14, Applicati  | 1.16e-02  |
| 9          | 51    | 54.3        | 295    | 2     | US-08-504- Sequence 12, Applicati  | 1.16e-02  |
| 10         | 51    | 54.3        | 297    | 2     | US-08-504- Sequence 12, Applicati  | 1.16e-02  |
| 11         | 51    | 54.3        | 625    | 1     | US-08-365- Sequence 13, Applicatio | 1.45e-02  |
| 12         | 50    | 53.2        | 316    | 2     | US-08-846- Sequence 2, Applicatio  | 1.45e-02  |
| 13         | 50    | 53.2        | 317    | 3     | US-08-951- Sequence 2, Applicatio  | 1.45e-02  |
| 14         | 50    | 53.2        | 317    | 2     | US-09-066- Sequence 2, Applicatio  | 1.45e-02  |
| 15         | 50    | 53.2        | 317    | 2     | US-08-518- Sequence 2, Applicatio  | 1.45e-02  |
| 16         | 50    | 53.2        | 908    | 2     | US-08-249- Sequence 2, Applicatio  | 1.45e-02  |
| 17         | 49    | 52.1        | 299    | 2     | US-08-504- Sequence 10, Applicati  | 1.81e-02  |
| 18         | 48    | 51.1        | 85     | 2     | US-08-659- Sequence 3, Applicatio  | 2.27e-02  |
| 19         | 48    | 51.1        | 85     | 2     | US-08-480- Sequence 3, Applicatio  | 2.27e-02  |
| 20         | 48    | 51.1        | 111    | 1     | US-08-111- Sequence 23, Applicati  | 2.27e-02  |
| 21         | 48    | 51.1        | 160    | 2     | US-08-162- Sequence 14, Applicati  | 2.27e-02  |
| 22         | 48    | 51.1        | 218    | 2     | US-08-162- Sequence 4, Applicatio  | 2.27e-02  |
| 23         | 48    | 51.1        | 238    | 1     | US-07-607- Sequence 4, Applicatio  | 2.27e-02  |

|    |    |      |      |   |                                   |          |
|----|----|------|------|---|-----------------------------------|----------|
| 24 | 48 | 51.1 | 357  | 1 | US-08-356- Sequence 2, Applicatio | 2.27e-02 |
| 25 | 48 | 51.1 | 357  | 2 | US-08-031- Sequence 4, Applicatio | 2.27e-02 |
| 26 | 48 | 51.1 | 400  | 2 | US-08-474- Sequence 14, Applicati | 2.27e-02 |
| 27 | 48 | 51.1 | 400  | 2 | US-08-118- Sequence 14, Applicati | 2.27e-02 |
| 28 | 48 | 51.1 | 400  | 2 | US-08-488- Sequence 14, Applicati | 2.27e-02 |
| 29 | 48 | 51.1 | 400  | 2 | US-08-486- Sequence 14, Applicati | 2.27e-02 |
| 30 | 48 | 51.1 | 407  | 2 | US-08-926- Sequence 2, Applicatio | 2.27e-02 |
| 31 | 48 | 51.1 | 407  | 3 | US-09-119- Sequence 2, Applicatio | 2.27e-02 |
| 32 | 48 | 51.1 | 432  | 2 | US-08-896- Sequence 5, Applicatio | 2.27e-02 |
| 33 | 48 | 51.1 | 432  | 2 | US-08-896- Sequence 4, Applicatio | 2.27e-02 |
| 34 | 48 | 51.1 | 863  | 1 | US-08-325- Sequence 4, Applicatio | 2.27e-02 |
| 35 | 48 | 51.1 | 1724 | 1 | US-08-325- Sequence 3, Applicatio | 2.27e-02 |
| 36 | 47 | 50.0 | 410  | 2 | US-08-471- Sequence 40, Applicati | 2.83e-02 |
| 37 | 47 | 50.0 | 410  | 2 | US-08-463- Sequence 43, Applicati | 2.83e-02 |
| 38 | 47 | 50.0 | 410  | 2 | US-08-471- Sequence 43, Applicati | 2.83e-02 |
| 39 | 47 | 50.0 | 410  | 1 | US-08-471- Sequence 43, Applicati | 2.83e-02 |
| 40 | 47 | 50.0 | 449  | 2 | US-08-463- Sequence 45, Applicati | 2.83e-02 |
| 41 | 47 | 50.0 | 449  | 2 | US-08-471- Sequence 45, Applicati | 2.83e-02 |
| 42 | 47 | 50.0 | 462  | 2 | US-08-471- Sequence 2, Applicatio | 2.83e-02 |
| 43 | 47 | 50.0 | 462  | 2 | US-08-471- Sequence 2, Applicatio | 2.83e-02 |
| 44 | 47 | 50.0 | 1338 | 2 | US-08-471- Sequence 50, Applicati | 2.83e-02 |
| 45 | 47 | 50.0 | 1346 | 1 | US-08-471- Sequence 21, Applicati | 2.83e-02 |

ALIGNMENTS

Sequence 4, Application US/08459264

Sequence 4, Application US/08459264

Patent No. 5705340

GENERAL INFORMATION:

APP INVENTOR: RASMUSSEN, Beth A

APP INVENTOR: TALLY, Francis P

APPLICANT: GLUZMAN, Yakov

TITLE OF INVENTION: DNA SEQUENCES AND AMINO ACID SEQUENCES

TITLE OF INVENTION: OF CLASS B BETA-LACTAMASE ENZYMES FROM BACTERIOIDES

NUMBER OF SEQUENCES: 9

CORRESPONDENCE ADDRESS:

ADDRESSEE: Darby & Darby PC

STREET: 805 Third Avenue

CITY: New York

STATE: NY

COUNTRY: US

ZIP: 10022

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08459,264

FILING DATE: 02-JUN-1995

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Robinson, Joseph R

| 33 | 448 | REGISTRATION NUMBER: 33,448 |
| 33 | 448 | REFERENCE/DOCKET NUMBER: 0646/1B026-US3 |
| 33 | 448 | TELECOMMUNICATION INFORMATION: |
| 33 | 448 | TELEPHONE: 212-527-7700 |
| 33 | 448 | TELEFAX: 212-753-6237 |
| 33 | 448 | TELEX: 236687 |
| 33 | 448 | INFORMATION FOR SEQ ID NO: 4: |
| 33 | 448 | SEQUENCE CHARACTERISTICS: |
| 33 | 448 | LENGTH: 255 amino acids |
| 33 | 448 | TYPE: amino acid |

STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: peptide  
 ORIGINAL SOURCE: Bacillus cereus  
 ORGANISM: Bacillus cereus  
 IMMEDIATE SOURCE: B-LACTAMASE  
 CLONE: B-LACTAMASE  
 SEQUENCE 255 AA: 27847 MW: 349604 CN:

Query Match 59.6% Score 56: DB 1: Length 255:  
 Best Local Similarity 70.0%: Pred. No. 3,66e-01:  
 Matches 7: Conservative 2: Mismatches 1: Indels 0: Gaps 0:

DB 103 FOKRVIDVII 112  
 I I I I I I I I  
 QY 2 FIKRVSNVII 11

RESULT 2  
 ID US-09-459-263-4 STANDARD: PRT: 255 AA:  
 XX  
 AC xxxxxx  
 XX  
 DT  
 XX

Sequence 4: Application US/04459263

Sequence 4: Application US/04459263

Patent No. 5793642

GENERAL INFORMATION:

APPLICANT: RASMUSSEN, Beth A

APPLICANT: TALSY, Francis P

APPLICANT: GLJZMAN, Yakov

TITLE OF INVENTION: CNA SEQUENCES AND AMING ACID SEQUENCES

TITLE OF INVENTION: FRAGILES

NUMBER OF SEQUENCES: 9

CORRESPONDENCE ADDRESS:

ADDRESSEE: Darby & Darby PC

STREET: 805 Third Avenue

CITY: New York

STATE: NY

COUNTRY: US

ZIP: 10022

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/04459263

FILING DATE: 02-JUN-1995

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Robinson, Joseph R

REGISTRATION NUMBER: 33,448

REFERENCE/DOCKET NUMBER: 0646/18026-US2

TELECOMMUNICATION INFORMATION:

TELEPHONE: 212-527-7700

TELEFAX: 212-753-6237

TELEX: 236687

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 255 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide

ORIGINAL SOURCE: Bacillus cereus

ORGANISM: Bacillus cereus

IMMEDIATE SOURCE: B-LACTAMASE

SEQUENCE 255 AA: 27847 MW: 349604 CN:

Query Match 59.6% Score 56: DB 1: Length 255:  
 Best Local Similarity 70.0%: Pred. No. 3,66e-01:  
 Matches 7: Conservative 2: Mismatches 1: Indels 0: Gaps 0:

DB 103 FOKRVIDVII 112  
 I I I I I I I I  
 QY 2 FIKRVSNVII 11

RESULT 3  
 ID US-09-120-353-2 STANDARD: PRT: 408 AA:  
 XX  
 AC xxxxxx  
 XX  
 DT  
 XX

Sequence 2: Application US/09120054

Sequence 2: Application US/09120054

Patent No. 5932454

GENERAL INFORMATION:

APPLICANT: Baltz, Richard H.

APPLICANT: Solenberg, Patricia J.

APPLICANT: Treadway, Patti J.

TITLE OF INVENTION: Glucosyltransferase Gene and Protein

TITLE OF INVENTION: Amycolatopsis orientalis

NUMBER OF SEQUENCES: 3

CORRESPONDENCE ADDRESS:

ADDRESSEE: Eli Lilly and Company

STREET: Lilly Corporate Center

CITY: Indianapolis

STATE: Indiana

COUNTRY: USA

ZIP: 46285

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09120054

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/946,258

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Webster, Thomas L.

REGISTRATION NUMBER: 4,471

REFERENCE/DOCKET NUMBER: 044,471

TELECOMMUNICATION INFORMATION:

TELEPHONE: 317-276-3334

TELEFAX: 317-276-3861

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 408 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE 408 AA: 43042 MW: 779745 CN:

Query Match 59.6% Score 56: DB 2: Length 408:  
 Best Local Similarity 53.8%: Pred. No. 3,66e-01:  
 Matches 7: Conservative 4: Mismatches 2: Indels 0: Gaps 0:

DB 298 VLFRRVAAVIHG 310  
 I I I I I I I I  
 QY 1 VFIKRVSNVIIH 13

RESULT 4  
 ID US-08-926-258-2 STANDARD: PRT: 408 AA:

```
XX AC xxxxxx
XX DT
XX XX
DE DE Sequence 2, Application US/08925258
XX XX
XX XX
CC CC Sequence 2, Application US/08925258
CC CC Patent No. 5871983
CC CC GENERAL INFORMATION:
CC CC APPLICANT: Balitz, Richard H.
CC CC APPLICANT: Solenberg, Patricia J.
CC CC APPLICANT: Ireadway, Patti J.
CC CC TITLE OF INVENTION: Glucosyltransferase Gene safe From
CC CC TITLE OF INVENTION: Atycolatopsis orientalis
CC CC NUMBER OF SEQUENCES: 3
CC CC CORRESPONDENCE ADDRESS:
CC CC ADDRESSEE: Eli Lilly and Company
CC CC STREET: Lilly Corporate Center
CC CC CITY: Indianapolis
CC CC STATE: Indiana
CC CC COUNTRY: USA
CC CC ZIP: 46285
CC CC COMPUTER READABLE FORM:
CC CC MEDIUM TYPE: floppy disk
CC CC COMPUTER: IBM PC compatible
CC CC OPERATING SYSTEM: PC-DOCS/MS-DOCS
CC CC SOFTWARE: Patent Release #1.0, Version #1.30
CC CC CURRENT APPLICATION DATA:
CC CC APPLICATION NUMBER: US/08/926,258
CC CC FILING DATE:
CC CC CLASSIFICATION: 435
CC CC ATTORNEY/AGENT INFORMATION:
CC CC NAME: Webster, Thomas D.
CC CC REGISTRATION NUMBER: 39,872
CC CC REFERENCE/DOCKET NUMBER: P-10430
CC CC TELECOMMUNICATION INFORMATION:
CC CC TELEPHONE: 317-276-3334
CC CC TELEFAX: 317-276-3861
CC CC INFORMATION FOR SEQ ID NO: 2:
CC CC SEQUENCE CHARACTERISTICS:
CC CC LENGTH: 408 amino acids
CC CC TYPE: amino acid
CC CC TOPOLOGY: linear
CC CC MCLEULE TYPE: protein
CC CC SEQUENCE 408 AA: 43042 MW: 779745 CN:

Query Match 59.6%; Score 56; DB 2; Length 408;
Best Local Similarity 53.8%; Pred. No. 3,66e-01;
Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

DE 298 YLFRRAVAIVHG 310
QY 1 VFIRKVSNIING 13
RESULT 5
ID PCT-US95-10398-80 STANDARD: PRI: 192 AA.
XX AC xxxxxx
XX DT
XX XX
DE DE Sequence 80, Application PC/TUS9510398
XX XX
XX XX
CC CC Sequence 80, Application PC/TUS9510398
CC CC GENERAL INFORMATION:
CC CC APPLICANT: BUKH, J., MILLER, R.H. AND
CC CC APPLICANT: PURCELL, R.H.
CC CC TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
CC CC TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
CC CC TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
CC CC TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
CC CC NUMBER OF SEQUENCES: 159
CC CC CORRESPONDENCE ADDRESS:

TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
NUMBER OF SEQUENCES: 263
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORGAN & FINNEGAN
STREET: 345 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10154
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOCS/MS-DOCS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/10398
FILING DATE: 15-AUG-1995
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/086,428
FILING DATE: 29 JUNE 1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/290,665
FILING DATE: 15 AUGUST 1994
ATTORNEY/AGENT INFORMATION:
NAME: RICHARD W. BORK
REGISTRATION NUMBER: 36,459
REFERENCE/DOCKET NUMBER: 2026-4116
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 758-4800
TELEFAX: (212) 751-6849
TELEX: 421792
INFORMATION FOR SEQ ID NO: 80:
SEQUENCE CHARACTERISTICS:
LENGTH: 192 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
ORIGINAL SOURCE:
ORGANISM: Homosapiens
INDIVIDUAL ISOLATE: 0
SEQUENCE 192 AA: 20855 MW: 90760 CN:

Query Match 55.3%; Score 52; DB 4; Length 192;
Best Local Similarity 50.0%; ed. No. 9,20e+01;
Matches 5; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

Db 145 YVMRVPEVII 154
QY 2 FIKRVSNVII 1;
RESULT 6
ID US-08-086-428B-80 STANDARD: PRI: 192 AA.
XX AC xxxxxx
XX DT
XX XX
DE DE Sequence 80, Application US/08086428B
XX XX
CC CC Sequence 80, Application US/08086428B
CC CC Patent No. 5514539
CC CC GENERAL INFORMATION:
CC CC APPLICANT: BUKH, J., MILLER, R.H. AND
CC CC APPLICANT: PURCELL, R.H.
CC CC TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
CC CC TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 GENE
CC CC TITLE OF INVENTION: OF 51 ISOLATES OF HEPATITIS C AND THE USE
CC CC TITLE OF INVENTION: OF REAGENTS DERIVED FROM THESE SEQUENCES IN
CC CC TITLE OF INVENTION: DIAGNOSTIC METHODS AND VACCINES
CC CC NUMBER OF SEQUENCES: 159
CC CC CORRESPONDENCE ADDRESS:
```

CC ADDRESSEE: MORGAN & FINNEGAN  
 CC STREET: 345 PARK AVENUE  
 CC CITY: NEW YORK  
 CC STATE: NEW YORK  
 CC COUNTRY: USA  
 CC ZIP: 10154  
 CC COMPUTER READABLE FORM:  
 CC MEDIUM TYPE: FLOPPY DISK  
 CC COMPUTER: IBM PC COMPATIBLE  
 CC OPERATING SYSTEM: PC-DOS/MS-DOS  
 CC SOFTWARE: WORDPERFECT 5.1  
 CC CURRENT APPLICATION DATA:  
 CC APPLICATION NUMBER: US/08/086.428B  
 CC FILING DATE: 29-JUN-1993  
 CC CLASSIFICATION: 424  
 CC ATTORNEY/AGENT INFORMATION:  
 CC NAME: RICHARD W. BORK  
 CC REGISTRATION NUMBER: 36.459  
 CC REFERENCE/DOCKET NUMBER: 2026-4270  
 CC TELECOMMUNICATION INFORMATION:  
 CC TELEPHONE: (212) 758-4800  
 CC TELEFAX: (212) 751-6849  
 CC TELEX: 421792  
 CC INFORMATION FOR SEQ ID NO: 80:  
 CC SEQUENCE CHARACTERISTICS:  
 CC LENGTH: 192 amino acids  
 CC TYPE: amino acid  
 CC STRANDEDNESS: unknown  
 CC TOPOLOGY: unknown  
 CC ORIGINAL SOURCE:  
 CC ORGANISM: homosapiens  
 CC INDIVIDUAL ISOLATE: US10  
 CC SEQUENCE: 192 AA; 20855 MW; 200760 CN;

Query Match: 55.3% Score 52: 08.2; Length 192;  
 Best Local Similarity 50.0%; Pred. No. 9.20e-01;  
 Matches 5; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

DB 145 YVNRVPEVIL 154  
 QY 2 FIKRVSNVII 11

PER-JUL 7  
 CC US-09-290-665A-60 STANDARD: PRT: 192 AA.  
 CC XXXXXX

DE Sequence 80, Application US/08/290665A  
 XX Sequence 80, Application US/08/290665A  
 CC Patent No. 5882552

CC GENERAL INFORMATION:  
 CC APPLICANT: BUKH, J., MILLER, R. H. AND  
 CC TITLE OF INVENTION: NUCLEOTIDE AND DECODED  
 CC TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND  
 CC TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS  
 CC TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE  
 CC TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES  
 CC NUMBER OF SEQUENCES: 263  
 CC CORRESPONDENCE ADDRESS:  
 CC ADDRESSEE: MORGAN & FINNEGAN  
 CC STREET: 345 PARK AVENUE  
 CC CITY: NEW YORK  
 CC STATE: NEW YORK  
 CC COUNTRY: USA  
 CC ZIP: 10154  
 CC COMPUTER READABLE FORM:  
 CC MEDIUM TYPE: FLOPPY DISK  
 CC COMPUTER: IBM PC COMPATIBLE

CC OPERATING SYSTEM: PC-DOS/MS-DOS  
 CC SOFTWARE: WORDPERFECT 5.1  
 CC CURRENT APPLICATION DATA:  
 CC APPLICATION NUMBER: 08/290.665A  
 CC FILING DATE: 15-AUG-1  
 CC CLASSIFICATION: 435  
 CC ATTORNEY/AGENT INFORMATION:  
 CC NAME: RICHARD W. BORK  
 CC REGISTRATION NUMBER: 36.454  
 CC REFERENCE/DOCKET NUMBER: 2026-4115  
 CC TELECOMMUNICATION INFORMATION:  
 CC TELEPHONE: (212) 758-4800  
 CC TELEFAX: (212) 751-6849  
 CC TELEX: 421792  
 CC INFORMATION FOR SEQ ID NO: 80:  
 CC SEQUENCE CHARACTERISTICS:  
 CC LENGTH: 192 amino acid  
 CC TYPE: amino acid  
 CC STRANDEDNESS: unknown  
 CC TOPOLOGY: unknown  
 CC ORIGINAL SOURCE:  
 CC ORGANISM: homosapiens  
 CC INDIVIDUAL ISOLATE: US10  
 CC SEQUENCE: 192 AA; 20855 MW; 200760 CN;

Query Match: 55.3% Score 52: 08.2; Length 192;  
 Best Local Similarity 50.0%; Pred. No. 9.20e-01;  
 Matches 5; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

DB 145 YVNRVPEVIL 154  
 QY 2 FIKRVSNVII 11

RESULT 2  
 ID US-08-468-570-80 STANDARD: PRT: 192 AA.  
 XX XXXXXX

Sequence 80, Application US/08468570

Sequence 80, Application US/ 8458570  
 Patent No. 5873562

CC GENERAL INFORMATION:  
 CC APPLICANT: BUKH, J., MILLER, R. H. AND  
 CC TITLE OF INVENTION: NUCLEOTIDE AND DECODED  
 CC TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 GENE  
 CC TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C AND THE USE  
 CC TITLE OF INVENTION: OF REAGENTS DERIVED FROM THESE SEQUENCES IN  
 CC TITLE OF INVENTION: DIAGNOSTIC METHODS AND VACCINES  
 CC NUMBER OF SEQUENCES: 159

CC CORRESPONDENCE ADDRESS:  
 CC ADDRESSEE: MORGAN & FINNEGAN  
 CC STREET: 345 PARK AVENUE  
 CC CITY: NEW YORK  
 CC STATE: NEW YORK  
 CC COUNTRY: USA  
 CC ZIP: 10154

CC COMPUTER READABLE FORM:  
 CC MEDIUM TYPE: FLOPPY DISK  
 CC COMPUTER: IBM PC COMPATIBLE  
 CC OPERATING SYSTEM: PC-DOS/MS-DOS  
 CC SOFTWARE: WORDPERFECT 5.1  
 CC CURRENT APPLICATION DATA:  
 CC APPLICATION NUMBER: 08/468.570  
 CC FILING DATE: 6-JUN-93  
 CC CLASSIFICATION: 424  
 CC PRIOR APPLICATION DATA:  
 CC APPLICATION NUMBER: 08/086.428  
 CC FILING DATE: 29-JUN-1993

CC CLASSIFICATION: 424  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: RICHARD W. BORK  
CC REGISTRATION NUMBER: 36,459  
CC REFERENCE/DOCKET NUMBER: 2026-4070US1  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (212) 758-4800  
CC TELEFAX: (212) 751-6849  
CC TELEX: 421792  
CC INFORMATION FOR SEQ ID NO: 80:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 192 amino acids  
CC TYPE: amino acid  
CC STRANDEDNESS: unknown  
CC TOPOLOGY: unknown  
CC ORIGIN: SOURCE:  
CC ORGANISM: homosapiens  
CC INDIVIDUAL ISOLATE: US10  
CC SEQUENCE: 192 AA; 20855 MW; 200760 CN;

Query Match: 55.3%; Score 52; DB 2; Length 192;  
Best Local Similarity 50.0%; Pred. NO. 1.16e-01;  
Matches 5; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

DB 145 YVMRVPV11 154  
QY 2 FIKRVSNV11 11

RESULT 5  
ID US-08-504-459-14 STANDARD: PRT: 295 AA.

XX

AC xxxxxx

XX

DT

XX

Sequence 14, Application US/08504459

Sequence 14, Application US/08504459

Patent No. 5922563

GENERAL INFORMATION:

APPLICANT: Alderete, John F.

TITLE OF INVENTION: Adhesin Genes and Proteins Involved in

TITLE OF INVENTION: Trichomonas Vaginalis Cytoadherence

NUMBER OF SEQUENCES: 30

CORRESPONDENCE ADDRESS:

ADDRESSEE: Arnold, White & Durkee

STREET: P.O. Box 4433

CITY: Houston

STATE: Texas

COUNTRY: US

ZIP: 77210

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/504,459

FILING DATE: Concurrently herewith

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Wilson, Mark B.

REGISTRATION NUMBER: 37,259

REFERENCE/DOCKET NUMBER: US/08/504,459

TELECOMMUNICATION INFORMATION:

TELEPHONE: 512/418-3000

TELEFAX: 512/474-7577

TELEX: N/A

INFORMATION FOR SEQ ID NO: 14:

SEQUENCE CHARACTERISTICS:

LENGTH: 295 amino acids

TYPE: amino acid

CC TOPOLOGY: linear  
CC MOLECULE TYPE: protein  
SQ SEQUENCE 295 AA; 30887 MW; 423730 CN;  
  
Query Match: 54.3%; Score 51; DB 2; Length 295;  
Best Local Similarity 38.5%; Pred. NO. 1.16e-02;  
Matches 5; Conservative 6; Mismatches 2; Indels 0; Gaps 0;  
  
DB 1 LFIDKDTKVV10G 13  
QY 1 VFIRKRVSNV11G 13  
  
RESULT 10  
ID US-08-504-459-12 STANDARD: PRT: 297 AA.  
XX  
AC xxxxxx  
XX  
DT  
XX  
XX  
DE Sequence 12, Application US/08504459  
XX  
CC Sequence 12, Application US/08504459  
CC Patent No. 5922563  
CC GENERAL INFORMATION:  
CC APPLICANT: Alderete, John F.  
CC TITLE OF INVENTION: Adhesin Genes and Proteins Involved in  
CC TITLE OF INVENTION: Trichomonas Vaginalis Cytoadherence  
CC NUMBER OF SEQUENCES: 30  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Arnold, White & Durkee  
CC STREET: P.O. Box 4433  
CC CITY: Houston  
CC STATE: Texas  
CC COUNTRY: US  
CC ZIP: 77210  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: PatentIn Release #1.0, Version #1.30  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/504,459  
CC FILING DATE: Concurrently herewith  
CC CLASSIFICATION: 435  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Wilson, Mark B.  
CC REGISTRATION NUMBER: 37,259  
CC REFERENCE/DOCKET NUMBER: US/08/504,459  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: 512/418-3000  
CC TELEFAX: 512/474-7577  
CC TELEX: N/A  
CC INFORMATION FOR SEQ ID NO: 12:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 297 amino acids  
CC TYPE: amino acid  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: protein  
SQ SEQUENCE 297 AA; 31082 MW; 429472 CN;

Query Match: 54.3%; Score 51; DB 2; Length 297;  
Best Local Similarity 38.5%; Pred. NO. 1.16e-02;  
Matches 5; Conservative 6; Mismatches 2; Indels 0; Gaps 0;

DB 3 LFIDKDTKVV10G 15  
QY 1 VFIRKRVSNV11G 13

RESULT 11

ID US-08-365-981-13 STANDARD: PRT: 625 AA.

XX





\*\*\*\*\*  
WORLD  
\*\*\*\*\*  
(TM)

Please J. A. John F. Collins, BioComputing Research Unit,  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MParch\_PP protein - protein database search, using Smith-Waterman algorithm  
Run on: Mon Jun 19 16:28:19 2000; Maspar time 14.74 Seconds  
Tabular output not generated. 88.845 Million cell updates/sec

Title: >US-09-142-524A-13  
Description: (1-13) from US09142524A.pep  
Perfect Score: 94  
Sequence: 1 VFIKRVSNVHIG 13

Scoring table: PAM 150  
Gap 15

Searched: 721208 seqs, 100765575 residues

Post-processing: Minimum Match 08  
Listing first 45 summaries

Database: a-pending  
1:PCI 2:JG 3:U60 4:U7 5:U80 6:U81 7:U82 8:U83 9:U84A  
10:U84B 11:U85 12:U86 13:U87 14:U88 15:U89 16:U90 17:U91  
18:U92 19:U93 20:U94 21:U95 22:NEWP 23:NEWU60 24:NEWJ8  
25:NEWJ9

Statistics: Mean: 20.556; Variance 54.778; scale 0.375

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description                       | Pred. No. |
|------------|-------|-------------|--------|----|-----------------------------------|-----------|
| 1          | 94    | 100.0       | 20     | 8  | US-08-350- Sequence 36, Applicati | 3.61e-03  |
| 2          | 94    | 100.0       | 20     | 7  | US-08-226- Sequence 36, Applicati | 3.61e-03  |
| 3          | 94    | 100.0       | 20     | 10 | US-08-467- Sequence 36, Applicati | 3.61e-03  |
| 4          | 94    | 100.0       | 20     | 4  | US-07-938- Sequence 36, Applicati | 3.61e-03  |
| 5          | 94    | 100.0       | 20     | 10 | US-08-468- Sequence 36, Applicati | 3.61e-03  |
| 6          | 94    | 100.0       | 20     | 10 | US-08-467- Sequence 36, Applicati | 3.61e-03  |
| 7          | 94    | 100.0       | 20     | 10 | US-08-467- Sequence 36, Applicati | 3.61e-03  |
| 8          | 94    | 100.0       | 20     | 8  | US-08-350- Sequence 62, Applicati | 3.61e-03  |
| 9          | 94    | 100.0       | 60     | 7  | US-08-226- Sequence 62, Applicati | 3.61e-03  |
| 10         | 94    | 100.0       | 60     | 10 | US-08-468- Sequence 62, Applicati | 3.61e-03  |
| 11         | 94    | 100.0       | 60     | 10 | US-08-467- Sequence 62, Applicati | 3.61e-03  |
| 12         | 94    | 100.0       | 60     | 4  | US-07-938- Sequence 62, Applicati | 3.61e-03  |
| 13         | 94    | 100.0       | 60     | 10 | US-08-467- Sequence 62, Applicati | 3.61e-03  |
| 14         | 94    | 100.0       | 60     | 10 | US-08-467- Sequence 62, Applicati | 3.61e-03  |
| 15         | 94    | 100.0       | 80     | 17 | US-09-142- Sequence 1, Applicatio | 3.61e-03  |
| 16         | 94    | 100.0       | 105    | 17 | US-09-142- Sequence 2, Applicatio | 3.61e-03  |
| 17         | 94    | 100.0       | 134    | 17 | US-09-142- Sequence 3, Applicatio | 3.61e-03  |
| 18         | 94    | 100.0       | 353    | 11 | US-08-526- Sequence 15, Applicati | 3.61e-03  |
| 19         | 94    | 100.0       | 374    | 4  | US-07-938- Sequence 2, Applicatio | 3.61e-03  |
| 20         | 94    | 100.0       | 374    | 10 | US-08-467- Sequence 2, Applicatio | 3.61e-03  |

|    |    |       |     |    |                                   |          |
|----|----|-------|-----|----|-----------------------------------|----------|
| 21 | 94 | 100.0 | 374 | 7  | US-08-226- Sequence 2, Applicatio | 3.61e-03 |
| 22 | 94 | 100.0 | 374 | 10 | US-08-467- Sequence 2, Applicatio | 3.61e-03 |
| 23 | 94 | 100.0 | 374 | 4  | US-07-930- Sequence 2, Applicatio | 3.61e-03 |
| 24 | 94 | 100.0 | 374 | 4  | 07-729- Sequence 2, Applicatio    | 3.61e-03 |
| 25 | 94 | 100.0 | 374 | 10 | 08-458- Sequence 2, Applicatio    | 3.61e-03 |
| 26 | 94 | 100.0 | 374 | 8  | 08-350- Sequence 2, Applicatio    | 3.61e-03 |
| 27 | 94 | 100.0 | 374 | 10 | US-08-467- Sequence 2, Applicatio | 3.61e-03 |
| 28 | 69 | 73.4  | 20  | 7  | US-08-226- Sequence 37, Applicati | 5.84e-00 |
| 29 | 69 | 73.4  | 20  | 8  | US-08-350- Sequence 37, Applicati | 5.84e-00 |
| 30 | 69 | 73.4  | 20  | 8  | US-08-467- Sequence 37, Applicati | 5.84e-00 |
| 31 | 69 | 73.4  | 20  | 10 | US-08-467- Sequence 37, Applicati | 5.84e-00 |
| 32 | 69 | 73.4  | 20  | 10 | US-08-467- Sequence 37, Applicati | 5.84e-00 |
| 33 | 69 | 73.4  | 20  | 4  | US-07-938- Sequence 37, Applicati | 5.84e-00 |
| 34 | 69 | 73.4  | 20  | 4  | US-07-938- Sequence 37, Applicati | 5.84e-00 |
| 35 | 67 | 71.3  | 370 | 8  | US-08-350- Sequence 97, Applicati | 1.02e-01 |
| 36 | 67 | 71.3  | 370 | 7  | US-08-226- Sequence 97, Applicati | 1.02e-01 |
| 37 | 67 | 71.3  | 370 | 10 | US-08-467- Sequence 97, Applicati | 1.02e-01 |
| 38 | 67 | 71.3  | 370 | 10 | US-08-467- Sequence 97, Applicati | 1.02e-01 |
| 39 | 67 | 71.3  | 370 | 10 | US-08-467- Sequence 97, Applicati | 1.02e-01 |
| 40 | 67 | 71.3  | 370 | 10 | US-08-467- Sequence 97, Applicati | 1.02e-01 |
| 41 | 64 | 68.1  | 31  | 17 | US-09-142- Sequence 5, Applicatio | 2.36e-01 |
| 42 | 64 | 68.1  | 31  | 17 | US-09-142- Sequence 4, Applicatio | 2.36e-01 |
| 43 | 61 | 64.9  | 367 | 8  | US-08-350- Sequence 95, Applicati | 5.36e-01 |
| 44 | 61 | 64.9  | 367 | 10 | US-08-467- Sequence 95, Applicati | 5.36e-01 |
| 45 | 61 | 64.9  | 367 | 10 | US-08-467- Sequence 95, Applicati | 5.36e-01 |

ALIGNMENTS

RESULT US-08-350-225-36 STANDARD: PMT: 20 AA.  
XX AC xxxxxx

Sequence 36, Application US/08350225

Sequence 36, Application US/08350225

GENERAL INFORMATION:

APPLICANT: Griffith, Irwin J.

APPLICANT: Pollock, Joanne

APPLICANT: Bond, Julia

APPLICANT: Garman, Richard D.

APPLICANT: Kuo, Mei-Chang

APPLICANT: Yeung, Siu-mei H.

APPLICANT: Brauer, And W.

APPLICANT: Exley, Mark A.

APPLICANT: Powers, Ste G. P.

TITLE OF INVENTION: A Genetic Proteins And Peptides From

NUMBER OF INVENTION: Japanese Cedar Pollen

CORRESPONDENCE ADDRESS:

ADDRESS: Immunologic Pharmaceutical Corporation, Inc.

STREET: 610 Lincoln St

CITY: Waltham

STATE: MA

COUNTRY: USA

ZIP: 02154

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/350,225

FILING DATE: December 6, 1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 07/226,248

FILING DATE: April 8, 1994

APPLICATION NUMBER: 07/938,990

FILING DATE: September 1, 1992



CC APPLICATION NUMBER: PCT/US93/00139  
CC FILING DATE: January 15, 1993  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Darlene A. Vanstone  
CC REGISTRATION NUMBER: 35,729  
CC REFERENCE/DOCKET NUMBER: 025.6 US (IM-0280CP)  
CC TELEPHONE: (617) 466-6000  
CC TELEFAX: (617) 466-6040  
CC INFORMATION FOR SEQ ID NO: 36:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 20 amino acids  
CC TYPE: amino acid  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: peptide  
CC FRAGMENT TYPE: internal  
CC SEQUENCE 20 AA: 2080 MW: 2304 DA:  
Query Match: 100.0% Score 94: 15.8 Length 20:  
Best Local Similarity 100.0% Pred. No. Mismatches 0: Indels 0: Gaps 0:  
Matches 13: Conservative 0: Mismatches 0: Indels 0: Gaps 0:  
DB 8 VFIRKRVSNV:HG 20  
QY 1 VFIRKRVSNV:HG 13  
RESULT 2  
ID US-08-226-248A-J6 STANDARD: PRI: 20 AA  
XX XXXXXX  
DE Sequence 36, Application US/08226248A  
CC Sequence 36, Application US/08226248A  
CC GENERAL INFORMATION:  
CC APPLICANT: Griffith, Irwin J.  
CC APPLICANT: Pollock, Joanne  
CC APPLICANT: Bond, Julian F.  
CC APPLICANT: Garman, Richard D.  
CC APPLICANT: Kuo, Mei-Chang  
CC APPLICANT: Yeung, Siu-mei H.  
CC APPLICANT: Brauer, Andrew  
CC APPLICANT: Exley, Mark A.  
CC APPLICANT: Powers, Steven P.  
CC TITLE OF INVENTION: Allergic Proteins And Peptides From  
CC TITLE OF INVENTION: Japanese Cedar Pollen  
CC NUMBER OF SEQUENCES: 20  
CC ADDRESS: Immunologic Pharmaceuticals Corporation, Inc  
CC STREET: 410 Lincoln St  
CC CITY: Waltham  
CC STATE: MA  
CC COUNTRY: USA  
CC ZIP: 02154  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: PatentIn Release #1.0, Version #1.25  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/226,248A  
CC FILING DATE: April 8, 1994  
CC CLASSIFICATION: 424  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: 07/938,990  
CC FILING DATE: September 1, 1992  
CC APPLICATION NUMBER: PCT/US93/00139  
CC FILING DATE: January 15, 1993  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Darlene A. Vanstone

CC REGISTRATION NUMBER: 35,729  
CC REFERENCE/DOCKET NUMBER: 025.6 US (IM-0280CP)  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (617) 466-6000  
CC TELEFAX: (617) 466-6040  
CC INFORMATION FOR SEQ ID NO: 36:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 20 amino acids  
CC TYPE: amino acid  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: peptide  
CC FRAGMENT TYPE: internal  
CC SEQUENCE 20 AA: 2080 MW: 2304 DA:  
Query Match: 100.0% Score 94: 15.8 Length 20:  
Best Local Similarity 100.0% Pred. No. Mismatches 0: Indels 0: Gaps 0:  
Matches 13: Conservative 0: Mismatches 0: Indels 0: Gaps 0:  
DB 8 VFIRKRVSNV:HG 20  
QY 1 VFIRKRVSNV:HG 13  
RESULT 3  
ID US-09-467-006-J6 STANDARD: PRI: 20 AA  
XX XXXXXX  
DE Sequence 36, Application US/467006  
CC Sequence 36, Application US/467006  
CC GENERAL INFORMATION:  
CC APPLICANT: Griffith, Irwin J.  
CC APPLICANT: Pollock, Joanne  
CC APPLICANT: Bond, Julian F.  
CC APPLICANT: Garman, Richard D.  
CC APPLICANT: Kuo, Mei-Chang  
CC APPLICANT: Yeung, Siu-mei H.  
CC APPLICANT: Brauer, Andrew  
CC APPLICANT: Exley, Mark A.  
CC APPLICANT: Powers, Steven P.  
CC TITLE OF INVENTION: Allergic Proteins And Peptides From  
CC TITLE OF INVENTION: Japanese Cedar Pollen  
CC NUMBER OF SEQUENCES: 261  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Immunologic Pharmaceuticals Corporation, Inc  
CC STREET: 410 Lincoln St  
CC CITY: Waltham  
CC STATE: MA  
CC COUNTRY: USA  
CC ZIP: 02154  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: PatentIn Release #1.0, Version #1.25  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/467,006  
CC FILING DATE: June 6, 1995  
CC CLASSIFICATION: 424  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: 08/350,225  
CC FILING DATE: December 6, 1994  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Jane E. Remillard  
CC REGISTRATION NUMBER: 38,872  
CC REFERENCE/DOCKET NUMBER: 025.6 US06 (IM-0280CPD6)  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (617) 227-7400  
CC TELEFAX: (617) 227-5941  
CC INFORMATION FOR SEQ ID NO: 36:

CC SEQUENCE CHARACTERISTICS:  
 CC LENGTH: 20 amino acids  
 CC TYPE: amino acid  
 CC TOPOLOGY: linear  
 CC MOLECULE TYPE: peptide  
 CC FRAGMENT TYPE: internal  
 CC SEQUENCE 20 AA: 2080 MW: 2304 CN:

Query Match 100.0% Score 94: DB 10: Length 20:  
 Best Local Similarity 100.0% Pred. No. 3.61e-03:  
 Matches 13: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

DB 8 VFIRKRVSNVING 20  
 QY 1 VFIRKRVSNVING 13

RESULT 4  
 ID US-07-938-992A-36 STANDARD: PRT: 20 AA.  
 AC xxxxxx

Sequence 36, Application US/07938992A

Sequence 36, Application US/07938992A

GENERAL INFORMATION:  
 CC APPLICANT: Griffith, Irwin J.  
 CC APPLICANT: Pollock, Joanne  
 CC APPLICANT: Bond, Julian F.  
 CC APPLICANT: Kuo, Mei-Chang  
 CC APPLICANT: Yeung, Shu-wei H.  
 CC APPLICANT: Brauer, Andrew  
 CC APPLICANT: Exley, Mark A.  
 CC APPLICANT: Powers, Steven P.  
 CC TITLE OF INVENTION: Allergenic Proteins And Peptides From  
 CC JAPANES CEDAR POLLEN  
 CC NUMBER OF SEQUENCES: 70

CC CORRESPONDENCE ADDRESS:  
 CC ADDRESSEE: Lahive & Cockfield  
 CC STREET: Sixty State Street  
 CC CITY: Boston  
 CC STATE: MA  
 CC COUNTRY: USA  
 CC ZIP: 02109

CC COMPUTER READABLE FORM:  
 CC MEDIUM TYPE: Floppy disk  
 CC OPERATING SYSTEM: IBM PC compatible  
 CC SOFTWARE: Patent In Release #1.0, Version #1.25  
 CC CURRENT APPLICATION DATA:  
 CC APPLICATION NUMBER: US/07/938-992A  
 CC FILING DATE: 19920901

CC CLASSIFICATION: 4.15

CC PRIORITY APPLICATION DATA:

CC APPLICATION NUMBER: 07/730,452

CC FILING DATE: July 15, 1991

CC PRIORITY APPLICATION DATA:

CC APPLICATION NUMBER: 07/729,134

CC FILING DATE: July 10, 1991

CC ATTORNEY/AGENT INFORMATION:

CC NAME: Amy E. Mandragoras

CC REGISTRATION NUMBER: 36,207

CC REFERENCE/DOCKET NUMBER: IPC-025CC (IMI-028)

CC TELECOMMUNICATION INFORMATION:

CC TELEPHONE: (617) 227-7400

CC TELEFAX: (617) 227-5941

CC INFORMATION FOR SEQ ID NO: 36:

CC SEQUENCE CHARACTERISTICS:

CC LENGTH: 20 amino acids

CC TYPE: AMINO ACID

CC TOPOLOGY: linear

CC MOLECULE TYPE: peptide

CC FRAGMENT TYPE: internal

CC SEQUENCE 20 AA: 2080 MW: 2304 CN:

Query Match 100.0% Score 94: DB 4: Length 20:

Best Local Similarity 100.0% Pred. No. 3.61e-03:  
 Matches 13: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

DB 8 VFIRKRVSNVING 20  
 QY 1 VFIRKRVSNVING 13

RESULT 5  
 ID US-08-468-940-36 STANDARD: PRT: 20 AA.  
 AC xxxxxx

Sequence 36, Application US/08468940

Sequence 36, Application US/08468940

GENERAL INFORMATION:  
 CC APPLICANT: Griffith, Irwin J.  
 CC APPLICANT: Pollock, Joanne  
 CC APPLICANT: Bond, Julian F.  
 CC APPLICANT: Garman, Richard D.  
 CC APPLICANT: Kuo, Mei-Chang  
 CC APPLICANT: Yeung, Shu-wei H.  
 CC APPLICANT: Brauer, Andrew  
 CC APPLICANT: Exley, Mark A.  
 CC APPLICANT: Powers, Steven P.  
 CC TITLE OF INVENTION: Allergenic Proteins And Peptides From  
 CC JAPANES CEDAR POLLEN  
 CC NUMBER OF SEQUENCES: 261

CC CORRESPONDENCE ADDRESS:  
 CC ADDRESSEE: Immunologic Pharmaceuticals Corporation, Inc.  
 CC STREET: 610 Lincoln St  
 CC CITY: Waltham  
 CC STATE: MA  
 CC COUNTRY: USA  
 CC ZIP: 02154

CC COMPUTER READABLE FORM:  
 CC MEDIUM TYPE: Floppy disk  
 CC OPERATING SYSTEM: IBM PC compatible  
 CC SOFTWARE: Patent In Release #1.0, Version #1.25  
 CC CURRENT APPLICATION DATA:  
 CC APPLICATION NUMBER: US/08/468-940  
 CC FILING DATE:

CC CLASSIFICATION: 4.24

CC PRIORITY APPLICATION DATA:

CC APPLICATION NUMBER: 08/350,245

CC FILING DATE: December 5, 1994

CC APPLICATION NUMBER: 08/226,149

CC FILING DATE: April 8, 1994

CC APPLICATION NUMBER: 07/938-990

CC FILING DATE: September 1, 1992

CC APPLICATION NUMBER: PCI/US93/00139

CC FILING DATE: January 15, 1993

CC ATTORNEY/AGENT INFORMATION:

CC NAME: Darlene A. Vanstone

CC REGISTRATION NUMBER: 35,729

CC REFERENCE/DOCKET NUMBER: 025.6 US (IMI-028CP2)

CC TELECOMMUNICATION INFORMATION:

CC TELEPHONE: (617) 466-6000

CC TELEFAX: (617) 466-6040

CC INFORMATION FOR SEQ ID NO: 36:

CC SEQUENCE CHARACTERISTIC:

CC LENGTH: 20 amino acids

CC TYPE: amino acid

CC TOPOLOGY: linear

CC MOLECULE TYPE: peptide

CC FRAGMENT TYPE: internal

CC SEQUENCE 20 AA: 2080 MW: 2304 CN:

Query Match 100.0% Score 94: DB 10: Length 20:

Best Local Similarity 100.0%; Pred. No. 3.61e-03;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 5 VFIKRVSNVING 20  
II IIIIIIIIIII  
QY 1 VFIKRVSNVING 13

RESULT 6  
ID US-08-467-697-36 STANDARD: PRT: 20 AA.  
XX  
AC xxxxxx

Sequence 36, Application US/08467697

Sequence 36, Application US/08467697  
GENERAL INFORMATION:

APPLICANT: Griffith, Irwin J.  
APPLICANT: Pollock, Joanne;  
APPLICANT: Bond, Julian F.  
APPLICANT: Garman, Richard D.  
APPLICANT: Kuo, Mei-Chang;  
APPLICANT: Yeung, Siu-mei H.  
APPLICANT: Brauer, Andrew;  
APPLICANT: Exley, Mark A.;  
APPLICANT: Powers, Steven P.  
APPLICANT: Griffith, Irwin J.  
TITLE OF INVENTION: Allergenic Proteins And Peptides From  
TITLE OF INVENTION: Japanese Cedar Pollen  
NUMBER OF SEQUENCES: 261  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Immunologic Pharmaceutical Corporation, Inc.  
STREET: 610 Lincoln St  
CITY: Waltham  
STATE: MA  
COUNTRY: USA  
ZIP: 02154

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/467,697  
FILING DATE: June 6, 1995  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/350,225  
FILING DATE: December 6, 1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Jane E. Remillard  
REGISTRATION NUMBER: 38,872  
REFERENCE/DOCKET NUMBER: 025.6 USD4 (IM1-028CPD4)  
TELEPHONE: (617) 227-7400  
TELEFAX: (617) 227-5941  
INFORMATION FOR SEQ ID NO: 36:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 20 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FRAGMENT TYPE: Internal  
SEQUENCE 20 AA: 2060 MW: 2304 CN:

Query Match 100.0%; Score 94; DB 10; Length 20;  
Best Local Similarity 100.0%; Pred. No. 3.61e-03;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 8 VFIKRVSNVING 20  
IIIIIIIIIIII  
QY 1 VFIKRVSNVING 13

RESULT 7  
ID US-08-467-023-36 STANDARD: PRT: 20 AA.  
XX  
AC xxxxxx

Sequence 36, Application US/08467023

Sequence 36, Application US/08467023  
GENERAL INFORMATION:

APPLICANT: Griffith, Irwin J.  
APPLICANT: Pollock, Joanne;  
APPLICANT: Bond, Julian F.  
APPLICANT: Garman, Richard D.  
APPLICANT: Kuo, Mei-Chang;  
APPLICANT: Yeung, Siu-mei H.  
APPLICANT: Brauer, Andrew;  
APPLICANT: Exley, Mark A.;  
APPLICANT: Powers, Steven P.  
APPLICANT: Griffith, Irwin J.  
TITLE OF INVENTION: Allergenic Proteins And Peptides From  
TITLE OF INVENTION: Japanese Cedar Pollen  
NUMBER OF SEQUENCES: 261  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Immunologic Pharmaceutical Corporation, Inc.  
STREET: 610 Lincoln St  
CITY: Waltham  
STATE: MA  
COUNTRY: USA  
ZIP: 02154

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/467,023  
FILING DATE: June 6, 1995  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/350,225  
FILING DATE: December 6, 1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Jane E. Remillard  
REGISTRATION NUMBER: 38,872  
REFERENCE/DOCKET NUMBER: 025.6 USD2 (IM1-028CPD2)  
TELEPHONE: (617) 227-7400  
TELEFAX: (617) 227-5941  
INFORMATION FOR SEQ ID NO: 36:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 20 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FRAGMENT TYPE: Internal  
SEQUENCE 20 AA: 2060 MW: 2304 CN:

Query Match 100.0%; Score 94; DB 10; Length 20;  
Best Local Similarity 100.0%; Pred. No. 3.61e-03;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 8 VFIKRVSNVING 20  
IIIIIIIIIIII  
QY 1 VFIKRVSNVING 13

RESULT 8  
ID US-08-350-225-62 STANDARD: PRT: 50 AA.  
XX  
AC xxxxxx

```

XX DE
XX DE
XX DE
XX DE
XX DE
Sequence 62, Application US/08350225
Sequence 62, Application US/08350225
GENERAL INFORMATION:
APPLICANT: Griffith, Irwin J.
APPLICANT: Pollock, Joanne
APPLICANT: Bond, Julian F.
APPLICANT: Garman, Richard D.
APPLICANT: Kuo, Mei-Chang
APPLICANT: Yeung, Siu-mei H.
APPLICANT: Brauer, Andrew
APPLICANT: Exley, Mark A.
APPLICANT: Powers, Steven P.
TITLE OF INVENTION: Allergenic Proteins And Peptides From
NUMBER OF SEQUENCES: 261
TITLE OF INVENTION: Japanese Cedar Pollen
CORRESPONDENCE ADDRESS:
ADDRESSEE: Immunologic Pharmaceutical Corporation, Inc.
STREET: 610 Lincoln St
CITY: Waltham
STATE: MA
COUNTRY: USA
ZIP: 02154
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08350225
FILING DATE: December 6, 1994
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/226,248
FILING DATE: April 8, 1994
APPLICATION NUMBER: 07/549,990
FILING DATE: September 1, 1992
APPLICATION NUMBER: PCT/US93/00139
FILING DATE: January 15, 1993
ATTORNEY/AGENT INFORMATION:
NAME: Darlene A. Vanstone
REGISTRATION NUMBER: 35,729
REFERENCE/DOCKET NUMBER: 025.6 US (IMI-028CP2)
TELEPHONE: (617) 466-6000
TELEFAX: (617) 466-6040
INFORMATION FOR SEQ ID NO: 62:
SEQUENCE CHARACTERISTICS:
LENGTH: 50 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: Peptide
FRAGMENT TYPE: Internal
SEQUENCE 60 AA: 6644 MW: 19464 CN:
Query Match 100.0% Score 94 DB 8 Length 60
Best Local Similarity 100.0% Pred. No. 3.61e-03
Matches 13: Conservative 0 Mismatches 0 Indels 0 Gaps 0
DB 48 VFIKRVSNVIHG 60
QY 1 VFIKRVSNVIHG 13
RESULT 9
ID US-08-226-248A-62 STANDARD: PRT: 60 AA.
XX xxxxxx
XX DE
XX DE
XX DE

```

```

XX DE
XX DE
XX DE
XX DE
XX DE
Sequence 62, Application US/08226248A
Sequence 62, Application US/08226248A
GENERAL INFORMATION:
APPLICANT: Griffith, Irwin J.
APPLICANT: Pollock, Joanne
APPLICANT: Bond, Julian F.
APPLICANT: Garman, Richard D.
APPLICANT: Kuo, Mei-Chang
APPLICANT: Yeung, Siu-mei H.
APPLICANT: Brauer, Andrew
APPLICANT: Exley, Mark A.
APPLICANT: Powers, Steven P.
TITLE OF INVENTION: Allergenic Proteins And Peptides From
NUMBER OF SEQUENCES: 261
TITLE OF INVENTION: Japanese Cedar Pollen
CORRESPONDENCE ADDRESS:
ADDRESSEE: Immunologic Pharmaceutical Corporation, Inc.
STREET: 610 Lincoln St
CITY: Waltham
STATE: MA
COUNTRY: USA
ZIP: 02154
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08226248A
FILING DATE: April 8, 1994
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/938,990
FILING DATE: September 1, 1992
APPLICATION NUMBER: PCT/US93/00139
FILING DATE: January 15, 1993
ATTORNEY/AGENT INFORMATION:
NAME: Darlene A. Vanstone
REGISTRATION NUMBER: 35,729
REFERENCE/DOCKET NUMBER: 025.5 US (IMI-028CP)
TELEPHONE: (617) 466-6000
TELEFAX: (617) 466-6040
INFORMATION FOR SEQ ID NO: 62:
SEQUENCE CHARACTERISTICS:
LENGTH: 60 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: Peptide
FRAGMENT TYPE: Internal
SEQUENCE 60 AA: 6644 MW: 19464 CN:
Query Match 100.0% Score 94 DB 7 Length 60
Best Local Similarity 100.0% Pred. No. 3.61e-03
Matches 13: Conservative 0 Mismatches 0 Indels 0 Gaps 0
DB 48 VFIKRVSNVIHG 60
QY 1 VFIKRVSNVIHG 13
RESULT 10
ID US-08-468-940-62 STANDARD: PRT: 60 AA.
XX xxxxxx
XX DE
XX DE
XX DE
Sequence 62, Application US/08468940
Sequence 62, Application US/08468940

```



CC APPLICANT: Young, Siu-mei H.;  
 CC APPLICANT: Brauer, Andrew;  
 CC APPLICANT: Exley, Mark A.;  
 CC APPLICANT: Powers, Steven P.  
 CC TITLE OF INVENTION: Allergenic Proteins And Peptides From  
 CC TITLE OF INVENTION: Japanese Cedar Pollen  
 CC NUMBER OF SEQUENCES: 261  
 CC CORRESPONDENCE ADDRESS:  
 CC ADDRESSEE: Immunologic Pharmaceutical Corporation, Inc.  
 CC STREET: 610 Lincoln St  
 CC CITY: Waltham  
 CC STATE: MA  
 CC COUNTRY: USA  
 CC ZIP: 02154  
 CC COMPUTER READABLE FORM:  
 CC MEDIUM TYPE: Floppy disk  
 CC OPERATING SYSTEM: PC-DOS/MS-DOS  
 CC SOFTWARE: Patent: Release #1.0, Version #1.25  
 CC CURRENT APPLICATION DATA:  
 CC APPLICATION NUMBER: US/09/467,697  
 CC FILING DATE: June 6, 1995  
 CC CLASSIFICATION: 424  
 CC PRIOR APPLICATION DATA:  
 CC APPLICATION NUMBER: 08/350,225  
 CC FILING DATE: December 6, 1994  
 CC ATTORNEY/AGENT INFORMATION:  
 CC NAME: Case E. Remillard  
 CC REGISTRATION NUMBER: 38,672  
 CC REFERENCE/DOCKET NUMBER: 025.6 USD4 (IMI-025CFD4)  
 CC TELECOMMUNICATION INFORMATION:  
 CC TELEPHONE: (617) 227-7430  
 CC TELEFAX: (617) 227-5941  
 CC INFORMATION FOR SEQ ID NO: 62:  
 CC SEQUENCE CHARACTERISTICS:  
 CC LENGTH: 60 amino acids  
 CC TYPE: amino acid  
 CC TOPOLOGY: linear  
 CC MOLECULE TYPE: Peptide  
 CC FRAGMENT TYPE: Internal  
 CC SEQUENCE 60 AA: 5644 MW: 19464 CN:  
 Query Match 100.0% Score 94: DB 10: Length 60:  
 Best Local Similarity 100.0% Pred. No. 3,61e-03:  
 Matches 13: Conservative 0: Mismatches 0: Indels 0: Gaps 0:  
 Db 48 VFIKRVSNVIHG 60  
 QY 1 VFIKRVSNVIHG 13  
 RESULT 13  
 ID US-07-938-990A-62 STANDARD: PRI: 60 AA.  
 XX xxxxxx  
 DE Sequence 62, Application US/07938990A  
 CC Sequence 62, Application US/07938990A  
 CC GENERAL INFORMATION:  
 CC APPLICANT: Griffith, Irwin J.  
 CC APPLICANT: Pollock, Joanne  
 CC APPLICANT: Bond, Julian  
 CC TITLE OF INVENTION: Allergenic Proteins And Peptides From  
 CC TITLE OF INVENTION: Japanese Cedar Pollen  
 CC NUMBER OF SEQUENCES: 70  
 CC CORRESPONDENCE ADDRESS:  
 CC ADDRESSEE: Lahive & Cockfield  
 CC STREET: Sixty State Street  
 CC CITY: Boston  
 CC STATE: MA

CC COUNTRY: USA  
 CC ZIP: 02109  
 CC COMPUTER READABLE FORM:  
 CC MEDIUM TYPE: Floppy disk  
 CC OPERATING SYSTEM: IBM PC compatible  
 CC SOFTWARE: Patent: Release #1.0, Version #1.25  
 CC CURRENT APPLICATION DATA:  
 CC APPLICATION NUMBER: 07/730,452  
 CC FILING DATE: July 15, 1991  
 CC CLASSIFICATION: 435  
 CC PRIOR APPLICATION DATA:  
 CC APPLICATION NUMBER: 07/729,134  
 CC FILING DATE: July 15, 1991  
 CC ATTORNEY/AGENT INFORMATION:  
 CC NAME: Amy E. Mandragoras  
 CC REGISTRATION NUMBER: 36,207  
 CC REFERENCE/DOCKET NUMBER: IPC-025CC (IMI-026)  
 CC TELECOMMUNICATION INFORMATION:  
 CC TELEPHONE: (617) 227-7400  
 CC TELEFAX: (617) 227-5411  
 CC INFORMATION FOR SEQ ID NO: 62:  
 CC SEQUENCE CHARACTERISTICS:  
 CC LENGTH: 60 amino acids  
 CC TYPE: AMINO ACID  
 CC TOPOLOGY: linear  
 CC MOLECULE TYPE: Peptide  
 CC FRAGMENT TYPE: Internal  
 CC SEQUENCE 60 AA: 6644 MW: 1464 CN:  
 Query Match 100.0% Score 94: DB 4: Length 60:  
 Best Local Similarity 100.0% Pred. No. 3,61e-03:  
 Matches 13: Conservative 0: Mismatches 0: Indels 0: Gaps 0:  
 Db 48 VFIKRVSNVIHG 60  
 QY 1 VFIKRVSNVIHG 13  
 RESULT 14  
 ID US-08-467-023-62 STANDARD: PRI: 60 AA.  
 XX xxxxxx  
 DE Sequence 62, Application US/08467023  
 CC Sequence 62, Application US/08467023  
 CC GENERAL INFORMATION:  
 CC APPLICANT: Griffith, Irwin J.  
 CC APPLICANT: Pollock, Joanne  
 CC APPLICANT: Bond, Julian  
 CC APPLICANT: Garman, Richard D.  
 CC APPLICANT: Kuc, Mei-Chang  
 CC APPLICANT: Yeung, Siu-mei H.  
 CC APPLICANT: Brauer, Andrew  
 CC APPLICANT: Exley, Mark A.  
 CC APPLICANT: Powers, Steven P.  
 CC TITLE OF INVENTION: Allergenic Proteins And Peptides From  
 CC TITLE OF INVENTION: Japanese Cedar Pollen  
 CC NUMBER OF SEQUENCES: 261  
 CC CORRESPONDENCE ADDRESS:  
 CC ADDRESSEE: Immunologic Pharmaceutical Corporation, Inc.  
 CC STREET: 610 Lincoln St  
 CC CITY: Waltham  
 CC STATE: MA  
 CC COUNTRY: USA  
 CC ZIP: 02154  
 CC COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/98/467,023  
 FILING DATE: June 6, 1995  
 CLASSIFICATION: 424  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 08/350,225  
 FILING DATE: December 6, 1994  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Jane E. Remillard  
 REGISTRATION NUMBER: 38,872  
 REFERENCE/DOCKET NUMBER: 025.5 USJ2 (IMI-02HCPD2)  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (617) 227-7400  
 TELEFAX: (617) 227-5941  
 INFORMATION FOR SEQ ID NO: 62:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 60 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: peptide  
 FRAGMENT TYPE: internal  
 SEQ SEQUENCE 60 AA: 6644 MW: 19464 CN:

Query Match 100.0% Score 94: DB 10: Length 60:  
 Best Local Similarity 100.0% Pred. No. 3.61e-03:  
 Matches 13: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

Db 48 VFIRKRVSNVILHG 60  
 QY 1 VFIRKRVSNVILHG 13

RESULT 15  
 ID US-09-142-524-1 STANDARD: PRT: 80 AA.  
 XX  
 AC xxxxxx  
 XX  
 DT  
 XX  
 DE  
 XX  
 XX  
 Sequence 1, Application US/09142524  
 Sequence 1, Application US/09142524  
 GENERAL INFORMATION:  
 APPLICANT: Sano, Toshio  
 APPLICANT: Kame, Akiohori  
 APPLICANT: Katsuki, Kazuo  
 APPLICANT: Ikeda, Akiko  
 APPLICANT: Kino, Kohsuke  
 TITLE OF INVENTION: Peptide-based Immunotherapeutic Agent For Treating  
 TITLE OF INVENTION: Allergic Diseases  
 FILE REFERENCE: Docket No. SPO-103  
 CURRENT APPLICATION NUMBER: US/09/142,524  
 CURRENT FILING DATE: 1999-01-04  
 EARLIER APPLICATION NUMBER: 8/80/702  
 EARLIER FILING DATE: 1996-03-10  
 EARLIER APPLICATION NUMBER: PCT/JP97/00740  
 EARLIER FILING DATE: 1997-03-10  
 NUMBER OF SEQ ID NOS: 5  
 SOFTWARE: PatentIn Ver. 2.0  
 SEQ ID NO 1  
 LENGTH: 80  
 TYPE: PRT  
 ORGANISM: Unknown  
 FEATURE:  
 OTHER INFORMATION: Description of Unknown Organism:peptide  
 SEQ SEQUENCE 80 AA: 9067 MW: 30888 CN:  
 Query Match 100.0% Score 94: DB 17: Length 80:  
 Best Local Similarity 100.0% Pred. No. 3.61e-03:

Matches 13: Conservative 0: Mismatches 0: Indels 0: Gaps 0:  
 Db 16 VFIRKRVSNVILHG 28  
 QY 1 VFIRKRVSNVILHG 13

Search completed: Mon Jun 19 16:28:35 2000  
 Job time : 16 secs.

\*\*\*\*\*

WATERMAN

(TM)

\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit,  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

WPSrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run On: Mon Jun 19 16:17:09 2000; Maspar time 5.14 Seconds

Tabular output not generated.  
119,270 Million cell updates/sec

Title: >US-09-142-524A-13  
Description: (1-13) from US09142524A.pep  
Perfect Score: 94  
Sequence: 1 VF1KRVSNV1:HG 13

Scoring table: PAW 150  
Gap 15

Searched: 142080 seqs, 47172406 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: pir62

1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 26.420; Variance 35.957; scale 0.735

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

| Result No. | Score | Query Match | ID   | Description | Pred. No.                      |
|------------|-------|-------------|------|-------------|--------------------------------|
| 1          | 94    | 100.0       | 374  | 2 J02124    | Major allergen Cry j 6.70e+08  |
| 2          | 94    | 100.0       | 374  | 2 J02123    | Major allergen Cry j 6.70e+08  |
| 3          | 61    | 64.9        | 326  | 2 S01509    | pectate lyase - Asper 9.65e+01 |
| 4          | 57    | 60.6        | 256  | 2 A32017    | Beta-lactamase (EC 3. 5.47e+00 |
| 5          | 57    | 60.6        | 300  | 2 F69719    | succinate--CoA ligase 5.47e+00 |
| 6          | 56    | 59.6        | 257  | 1 PNB82S    | Beta-lactamase (EC 3. 8.33e+00 |
| 7          | 56    | 59.6        | 257  | 1 PNB8U2    | Beta-lactamase (EC 3. 8.33e+00 |
| 8          | 55    | 58.5        | 317  | 2 JC5696    | prolyl aminopeptidase 1.26e+01 |
| 9          | 55    | 58.5        | 452  | 1 W2BE4     | gene 4 protein - huma 1.26e+01 |
| 10         | 55    | 58.5        | 550  | 2 T15868    | hypothetical protein 1.26e+01  |
| 11         | 55    | 58.5        | 1894 | 2 T02155    | DNA-directed DNA poly 1.26e+01 |
| 12         | 54    | 57.4        | 303  | 2 F70716    | probable succinyl-coa 1.90e+01 |
| 13         | 53    | 56.4        | 75   | 1 S74973    | virulence associated 2.85e+01  |
| 14         | 53    | 56.4        | 190  | 2 B2330     | transcription regulat 2.85e+01 |
| 15         | 53    | 56.4        | 317  | 2 G70356    | thymidylate synthase 2.85e+01  |
| 16         | 53    | 56.4        | 375  | 1 A28630    | muconate cycloisomera 2.85e+01 |
| 17         | 53    | 56.4        | 475  | 1 JN0327    | sulfate adenylyltrans 2.85e+01 |
| 18         | 53    | 56.4        | 835  | 1 JC6140    | cell surface-associat 2.85e+01 |
| 19         | 53    | 56.4        | 985  | 2 S99330    | Na+/H+-exchanging pro 2.85e+01 |
| 20         | 52    | 55.3        | 365  | 2 D75167    | hypothetical protein 4.25e+01  |
| 21         | 52    | 55.3        | 404  | 2 S12209    | pectate lyase (EC 4.2 4.25e+01 |
| 22         | 52    | 55.3        | 552  | 2 S46809    | hypothetical protein 4.25e+01  |
| 23         | 51    | 54.3        | 111  | 2 A72079    | hypothetical protein 6.30e+01  |

|    |    |      |      |          |                                 |
|----|----|------|------|----------|---------------------------------|
| 24 | 51 | 54.3 | 238  | 2 T16640 | hypothetical protein 6.30e+01   |
| 25 | 51 | 54.3 | 254  | 2 E71225 | hypothetical protein 6.30e+01   |
| 26 | 51 | 54.3 | 261  | 2 JN0871 | ribosomal protein S4 6.30e+01   |
| 27 | 51 | 54.3 | 379  | 2 S06193 | probable membrane pro 6.30e+01  |
| 28 | 51 | 54.3 | 379  | 2 S46187 | probable membrane pro 6.30e+01  |
| 29 | 51 | 54.3 | 381  | 2 S63322 | probable membrane pro 6.30e+01  |
| 30 | 51 | 54.3 | 720  | 2 S50920 | hypothetical protein 6.30e+01   |
| 31 | 51 | 54.3 | 785  | 2 B36847 | ATP/GTP-binding prote 6.30e+01  |
| 32 | 51 | 54.3 | 785  | 2 D42515 | D5R protein - vaccini 6.30e+01  |
| 33 | 51 | 54.3 | 785  | 1 Q0V28  | D5 protein - vaccinia 6.30e+01  |
| 34 | 51 | 54.3 | 810  | 1 I40508 | endopeptidase Clp ATP 6.30e+01  |
| 35 | 51 | 54.3 | 819  | 2 T13355 | hypothetical protein 6.30e+01   |
| 36 | 51 | 54.3 | 857  | 1 D35905 | endopeptidase Clp ATP 6.30e+01  |
| 37 | 51 | 54.3 | 1006 | 2 C70445 | ATPase subunit of ATP 6.30e+01  |
| 38 | 51 | 54.3 | 1161 | 2 B73172 | DNA polymerase III, s 6.30e+01  |
| 39 | 50 | 53.2 | 267  | 2 T08079 | carbonate dehydratase 9.27e+01  |
| 40 | 50 | 53.2 | 267  | 2 T08082 | carbonate dehydratase 9.27e+01  |
| 41 | 50 | 53.2 | 270  | 2 H64248 | hypothetical protein 9.27e+01   |
| 42 | 50 | 53.2 | 317  | 2 B74216 | endoglucanase - Therm 9.27e+01  |
| 43 | 50 | 53.2 | 403  | 2 JC5462 | endopolysialacturonase 9.27e+01 |
| 44 | 50 | 53.2 | 639  | 2 F71258 | probable heat shock p 9.27e+01  |
| 45 | 50 | 53.2 | 661  | 2 A64252 | 3-hydroxyacyl-CoA deh 9.27e+01  |

## ALIGNMENTS

RESULT 1  
ENTRY JC2124 #type complete  
TITLE major allergen Cry j : precursor (clone pCCI-15) - Japanese cedar:  
ORGANISM #formal\_name Cryptomeria japonica #common\_name Japanese cedar  
DATE 14-Jul-1994 #sequence\_revision 14-Jul-1994 #text\_change 26-Aug-1999  
ACCESSIONS JC2124  
REFERENCE JC2123  
#authors Sone, T.; Komiya, N.; Shimizu, K.; Kusakabe, T.; Morikubo, K.; Kino, K.  
#journal Biochem. Biophys. Res. Commun. (1994) 199:619-625  
#title Cloning and sequencing of cDNA coding for Cry j I, a major allergen of Japanese cedar pollen.  
#cross-references MIM:54183234  
#accession JC2124  
#molecule\_type mRNA  
##residues 1-374 ##label SON  
##cross-references GB:D26541; NID:9493633; PID:d1006097; PID:q493634  
##experimental\_source pollen  
##note the authors described carbohydrate binding site for:  
residue 279

CLASSIFICATION #superfamily for pectate lyase LA159  
KEYWORDS glycoprotein; pollen  
FEATURE  
1-2:  
22-374  
#domain signal sequence #status predicted #label SIG  
#product major allergen Cry j I (clone pCCI-15) #status predicted #label MATN  
158,191,293,354 #binding\_site carbohydrate (ASN) (covalent) #status predicted  
SUMMARY #length 374 #molecular-weight 40702 #checksum 3692  
Query Match 100.0%; Score 94; DB 2; Length 374;  
Best Local Similarity 100.0%; Pred. No. 6.70e+08;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 129 VF1KRVSNV1:HG 141  
|||||  
QY 1 VF1KRVSNV1:HG 13

RESULT 2  
ENTRY JC2123 #type complete  
TITLE major allergen Cry j I precursor (clone pCCI-2) - Japanese cedar  
ORGANISM #formal\_name Cryptomeria japonica #common\_name Japanese cedar  
DATE 14-Jul-1994 #sequence\_revision 14-Jul-1994 #text\_change





```

#title      The complete genome sequence of the Gram-positive bacterium
            Bacillus subtilis.
#cross-references GB:29112; GB:AA009126; NID:Q2633902;
#accession   F69719
#status      preliminary: nucleic acid sequence not shown;
            translation not shown
#molecule_type DNA
#residues    1-300
#authors     Hussain, M.; Carlino, A.; Madonna, M.J.; Lampen, J.O.
#cross-references GB:29112; GB:AA009126; NID:Q2633902;
            PIN:CA913483.1; PID:e185201; PID:Q2633982
#experimental_source strain 165
GENETICS
#gene        sscD
#classification superfamily succinate--CoA ligase (ADP-forming) alpha chain
#keywords    acid-thiol ligase; ATP; phosphohistidine; phosphoprotein;
            tricarboxylic acid cycle
FEATURE
247         #active_site His (phosphohistidine intermediate) #status
            predicted
SUMMARY      #length 300 #molecular_weight 3392 #checksum 2503
Query Match 60.6% Score 57; DB 2; Length 300;
Best Local Similarity 46.2% Pred. No. 5,47e-00;
Matches 6; Conservative 5; Mismatches 2; Indels 0; Gaps 0;
DB          3 VFINKIRIVVQ 15
          1111111111
QY          1 VFIKRVSNV11G 13
          1111111111
RESULT      6
ENTRY      PNB52S
#title     beta-lactamase (EC 3.5.2.6) II precursor - Bacillus sp.
#alternative_names penicillinase II
#organism   #formal_name Bacillus sp.
#date       31-Dec-1988 #sequence_revision 31-Dec-1988 #text_change
            18-Jun-1999
#accessions A24393; B24393
#reference   A24393
#authors     Kato, C.; Kudo, T.; Watanabe, K.; Horikoshi, K.
#journal     J. Gen. Microbiol. (1985) 131:3317-3324
#title       Nucleotide sequence of the beta-lactamase gene of
            alkalophilic Bacillus sp. strain 170.
#cross-references M1D:86170399
#accession   A24393
#molecule_type DNA
#residues    1-257
#experimental_source strain 170
#accession   B24393
#molecule_type protein
#residues    31-48
#classification superfamily beta-lactamase II
#keywords    antibiotic resistance; hydrolase; metalloprotein; zinc
            1-30
            #domain signal sequence #status predicted #label SIG\
31-257      #product beta-lactamase II #status predicted
67          #active_site Glu #status predicted\
116,118,198,240 #binding_site zinc, high affinity (His, His, Cys, His)
            #status experimental
SUMMARY      #length 257 #molecular_weight 28175 #checksum 2375
Query Match 59.6% Score 56; DB 1; Length 257;
Best Local Similarity 70.0% Pred. No. 8,33e-00;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
DB          105 FOKRVTDV11 114
          1111111111
QY          2 FIKRVSNV11 11
          1111111111
RESULT      7
ENTRY      PNB5U2
#title     type complete

```

```

TITLE      beta-lactamase (EC 3.5.2.6) II precursor - Bacillus cereus
ALTERNATE_NAMES cephalosporinase II; penicillinase II
ORGANISM        #formal_name Bacillus cereus
DATE            04-Dec-1995 #sequence_revision 04-Dec-1986 #text_change
            18-Jun-1999
#accessions A91806; A91344; A01006
#reference   A91806
#authors     Hussain, M.; Carlino, A.; Madonna, M.J.; Lampen, J.O.
#journal     J. Bacteriol. (1995) 154:223-229
#title       Cloning and sequencing of the metallothioesterase
            beta-lactamase II gene of Bacillus cereus 569/H in
            Escherichia coli.
#cross-references M1D:8508056
#accession   A91806
#molecule_type DNA
#residues    1-257
#cross-references GB:M1185; NID:Q142603; PID:AAA22276.1; PID:Q142604
#experimental_source strain 569/H
#reference   A91344
#authors     Ambler, R.P.; Daniele, M.; Fleming, J.; Hermoso, C.M.; Pand,
            C.; Waley, S.G.
#journal     FEBS Lett. (1985) 189:207-211
#title       The amino acid sequence of the zinc-requiring beta-lactamase
            II from the bacterium Bacillus cereus 569.
#cross-references M1D:8505445
#accession   A91344
#molecule_type protein
#residues    31-183; 187-210; 214-257
#experimental_source strain 569/H/H/
#comment     beta-lactamase II from Bacillus spp. and beta-lactamase L-1 from
            pseudomonas maltophilia are classified together as class B
            beta-lactamases, which hydrolyze cephalosporins and penicillins.
            beta-lactamase II binds two Zn(II) ions per molecule. Zinc at the
            higher affinity site is necessary for activity of the enzyme.
            Zinc at the lower affinity site (not specified) increases
            hydrolysis of cephalosporin C but not of benzylpenicillin.
#classification superfamily beta-lactamase II
#keywords     antibiotic resistance; hydrolase; metalloprotein; zinc
            1-30
            #domain signal sequence #status predicted #label SIG\
31-257      #product beta-lactamase II #status experimental #label
            MPI\
67          #active_site His #status predicted\
116,118,198,240 #binding_site zinc, high affinity (His, His, Cys, His)
            #status experimental
SUMMARY      #length 257 #molecular_weight 28092 #checksum 2889
Query Match 59.6% Score 55; DB 1; Length 257;
Best Local Similarity 70.0% Pred. No. 8,33e-00;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
DB          105 FOKRVTDV11 114
          1111111111
QY          2 FIKRVSNV11 11
          1111111111
RESULT      8
ENTRY      JC5696
#title     type complete
#alternative_names prolyl aminopeptidase (EC 3.4.11.5) - Serratia marcescens
            aminopeptidase; proline iminopeptidase
#organism     #formal_name Serratia marcescens
#date         20-Nov-1997 #sequence_revision 20-Nov-1997 #text_change
            29-Sep-1999
#accessions JC5696
#reference   JC5696
#authors     Kabashima, T.; Kitazono, A.; Kitano, A.; Ito, K.; Yoshimoto,
            T.
#journal     J. Biochem. (1997) 122:601-605
#title       Prolyl aminopeptidase from Serratia marcescens: Cloning of
            the enzyme gene and crystallization of the expressed
            enzyme.
#cross-references M1D:9806313

```

```

#accession JC5696
#molecule_type DNA
#residues 1-317 #label KAB
#cross-references DDBJ:U87897; NID:q2605614; PIDN:BAA23316.1;
#cross-references DDBJ:U87897; NID:q2605615
#comment This enzyme catalyzes the removal of amino-terminal proline from
peptides.

GENETICS
#gene spaf
#classification #superfamily proline aminopeptidase
#keywords alpha-aminocaproylpeptide hydrolase
#summary #length 317 #molecular-weight 35083 #checksum 7818

Query Match 58.5% Score 55: DB 2: Length 317:
Best Local Similarity 41.7% Pred. No. 1,26e-01:
Matches 5: Conservative 6: Mismatches 3: Indels 0: Gaps 0:

DB 254 LRRHIPAVVIRG 265
1 1111111111
2 FIKRVSNVIRG 13

RESULT 9
ENTRY WZBE4 #type complete
#description gene 4 protein - human herpesvirus 3
#formal_name human herpesvirus 3 varicella-zoster virus
#organism HCoV-1988 #sequence_revision 30-Sep-1988 #text_change
16-Jul-1999
#accession D27212
#molecule_type DNA
#residues 1-452 #label DAV
#cross-references EMBL:U04370; NID:q59989; PIDN:CAA27887.1; PID:q59993

GENETICS
#gene 4
#classification #superfamily varicella-zoster virus gene 4 protein
#keywords transcription regulation
#summary #length 452 #molecular-weight 51542 #checksum 254

Query Match 58.5% Score 55: DB 1: Length 452:
Best Local Similarity 38.5% Pred. No. 1,26e-01:
Matches 5: Conservative 6: Mismatches 2: Indels 0: Gaps 0:

DB 370 VMARIANVIRG 362
1 1111111111
2 FIKRVSNVIRG 13

RESULT 10
ENTRY T15868 #type complete
#description hypothetical protein C56G2.7 - Caenorhabditis elegans
#organism C. elegans
#date 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change
20-Sep-1999
#accessions T15868
#reference Z1842C
#authors Connel, M.
#submission submitted to the EMBL Data Library, March 1995
#description The sequence of C. elegans cosm.d C56G2.
#accession T15868
#status preliminary; translated from GB/EMBL/DBJ
#molecule_type DNA
#residues 1-550 #label CON
#cross-references EMBL:U23177; NID:q726411; PID:q726419;
PIDN:AAA64334.1; CESP:C56G2.7
#experimental_source strain Bristol N2

GENETICS
#gene CESP:C56G2.7

```

```

#introns 13/3: 70/3: 135/1: 165/1: 469/3: 525/1
#summary
#length 550 #molecular-weight 59367 #checksum 1654

Query Match 58.5% Score 55: DB 2: Length 550:
Best Local Similarity 50.0% Pred. No. 1,26e-01:
Matches 6: Conservative 6: Mismatches 3: Indels 0: Gaps 0:

DB 224 VFIKOSNDMLIH 235
1 1111111111
2 FIKRVSNVIRG 12

RESULT 11
ENTRY T02155 #type complete
#description DNA-directed DNA polymerase homolog - Arabidopsis thaliana
#formal_name Arabidopsis thaliana #molecule_type DNA
#organism Arabidopsis thaliana #sequence_revision 05-Mar-1999 #text_change
05-Mar-1999
#accessions T02155
#reference Z14592
#authors Vysotskaya, V.S.; Schwartz, J.R.; Toriumi, M.; Yan, G.; Li,
O.; Kwan, A.; Li, S.; Li, C.; Araceli, R.; Ag, M.; Brendel,
V.; Buehler, E.; Conway, A.B.; Conway, A.B.; Conway, A.B.;
Feng, J.; Kim, C.; Kurtz, D.; Li, Y.; Palm, C.; Salas,
P.; Sun, H.; Davis, R.W.; Ecker, J.R.; Federspiel, N.A.;
Theologis, A.
#submission submitted to the EMBL Data Library, June 1998
#description Arabidopsis thaliana chromosome 1 BAC T1F15 sequence
#accession T02155
#status preliminary; translated from GB/EMBL/DBJ
#molecule_type DNA
#residues 1-1894 #label VVS
#cross-references EMBL:AC004393; NID:q3180024; PID:q3176661;
GSPB:GN0059; AISP:T1F15.3

GENETICS
#gene AISP:T1F15.3
#map_position 1
#introns 38/1: 65/3: 85/1: 201/3: 141/3: 158/3: 181/3: 224/3: 291/1:
321/3: 642/3: 926/3: 1043/3: 1091/2: 1285/3: 1334/3:
1425/3: 1461/3: 1535/2: 1611/3: 1837/3
#summary #length 1894 #molecular-weight 211356 #checksum 9154

Query Match 58.5% Score 55: DB 2: Length 1894:
Best Local Similarity 50.0% Pred. No. 1,26e-01:
Matches 6: Conservative 6: Mismatches 3: Indels 0: Gaps 0:

DB 1693 YAEKRVVIRG 1704
1 1111111111
2 FIKRVSNVIRG 13

RESULT 12
ENTRY F70716 #type complete
#description probable succinyl-coa synthetase alpha chain - Mycobacterium
tuberculosis (strain H37Rv)
#organism Mycobacterium tuberculosis
#date 17-Jul-1998 #sequence_revision 17-Jul-1996 #text_change
16-Jul-1999
#accessions F70716
#reference A70500
#authors Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher,
C.; Harris, D.; Gordon, S.V.; Eiglmeier, K.; Gas, S.; Barry,
III, C.E.; Tekle, E.; Badcock, K.; Basham, D.; Brown, D.;
Chillingworth, T.; Connor, R.; Davies, K.; Devlin, K.;
Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.;
Hornsby, T.; Jacobs, K.; Krogh, A.; McLean, J.; Molle, S.;
Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.;
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.;
Skilton, S.; Squares, R.; Squires, R.; Sulston, J.E.;
Taylor, K.; Whitehead, S.; Barrell, B.G.
#journal Nature (1998) 393:537-544

```

```

#title      Deciphering the biology of Mycobacterium tuberculosis from
#            the complete genome sequence.
#cross-references MUID:98295987
#accession   F70716
##status     preliminary; nucleic acid sequence not shown;
#            translation not shown
##molecule_type DNA
##residues    1-303 ##label CCL
##cross-references GB:279730; GB:AL23456; NID:q1652627;
#            PID:CA801998.1; PID:e264228; PID:g1524210
##experimental_source strain H37Rv

GENETICS
#gene        sucD
#classification #superfamily succinate--CoA ligase (ADP-forming) alpha chain
#summary      #length 333 #molecular-weight 3229 #checksum 2091

Query Match 57.4% Score 54; DB 2; Length 303;
Best Local Similarity 38.5%; Pred. No. 1.90e+01;
Matches 5; Conservative 6; Mismatches 2; Indels 0; Gaps 0;

Db 6 IFLSRDNKVIIVQG 18
Qy 1 VFIRVSNVIIHG 13

RESULT 13
ENTRY S74973 #type complete
#title Virulence associated protein vaps - Synecocystis sp. (strain PCC 6803)
ALTERNATE_NAMES protein ssl2922
ORGANISM #formal_name Synecocystis sp.
#variety PCC 6803
#date 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
ACCESSIONS S74973
REFERENCE S74322
#authors Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.; Hirose, M.; Sugita, M.; Sasamoto, S.; Kimura, T.; Hosouchi, T.; Matsuno, A.; Muraki, A.; Nakazaki, N.; Naruo, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda, M.; Iabata, S.
#journal DNA Res. (1996) 3:109-136
#title Sequence analysis of the genome of the unicellular cyanobacterium Synecocystis sp. PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions.
#cross-references MUID:97061201
#accession S74973
#status     nucleic acid sequence not shown; translation not shown
##molecule_type DNA
##residues    1-75 ##label KAN
##cross-references EMBL:D90402; GB:AB001339; NID:q1652627;
#            PID:BAAL7013.1; PID:d1017746; PID:g1652088
##note       the nucleotide sequence was submitted to the EMBL Data Library, June 1996

GENETICS
#gene        vapB
#classification #superfamily virulence-associated protein vaps
#keywords     DNA binding; transcription regulation
#summary      #length 75 #molecular-weight 8539 #checksum 9185

Query Match 55.4% Score 53; DB 1; Length 75;
Best Local Similarity 45.5%; Pred. No. 2.85e+01;
Matches 5; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Db 28 VYIKKIGSTII 38
Qy 1 VFIRVSNVII 11

RESULT 14
ENTRY B72330 #type complete

```

```

#title      transcription regulator, TetR family - Thermotoga maritima
#            (strain MSB8)
#formal_name Thermotoga maritima
#date 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 20-Aug-1999
ACCESSIONS B72330
REFERENCE A72200
#authors Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwin, M.; Dodson, R.J.; Haft, D.; Hickey, E.K.; Peterson, S.D.; Nelson, W.C.; Ketchum, K.A.; McDonald, L.; Utterback, L.R.; Malek, J.A.; Linher, A.D.; Garrett, M.M.; Stewart, A.M.; Corbett, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.; Heidelberg, J.; Sutton, G.G.; Fleischmann, R.D.; White, O.; Salzberg, S.L.; Smith, W.O.; Venter, J.C.; Fraser, C.M.
#journal Nature (1999) 399:323-329
#title Evidence for lateral gene transfer between Archaea and Bacteria from genome sequence of Thermotoga maritima.
#cross-references MUID:99287316
#accession B72330
##status     preliminary
##molecule_type DNA
##residues    1-190 ##label ARN
##cross-references GB:AE001749; GB:AE005512; NID:q4981366; PID:g4981354;
#            IIGR:FM0823
##experimental_source strain MSB8

GENETICS
#gene        TM0823
#summary      #length 190 #molecular-weight 22452 #checksum 1501

Query Match 56.4% Score 53; DB 2; Length 190;
Best Local Similarity 58.3%; Pred. No. 2.85e+01;
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Db 126 FIKRVEELIEEG 137
Qy 2 FIKRVSNVIIHG 13

RESULT 15
ENTRY G70356 #type complete
#title Thymidylate synthase complementing proteins - Aquifex aeolicus
#formal_name Aquifex aeolicus
#date 08-May-1998 #sequence_revision 08-May-1998 #text_change 08-May-1998
ACCESSIONS G70356
REFERENCE A70356
#authors Derkx, C.; Wauters, P.V.; Baasland, J.; Young, W.G.; Tenck, A.L.; Zeman, D.B.; Overbeek, R.; Swad, W.A.; Keller, M.; Anjili, M.; Butler, K.; Feldman, R.A.; Shuff, J.M.; Olsen, G.J.; Swanson, K.V.
#journal Nature (1998) 392:355-358
#title The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
#cross-references MUID:98195666
#accession G70356
#status     preliminary; nucleic acid sequence not shown;
#            translation not shown
##molecule_type DNA
##residues    1-317 ##label AQF
##cross-references GB:AE00700; NID:q2983248; PID:q2983252; GB:AE00657
#            #experimental_source strain VF5

GENETICS
#gene        thy
#summary      #length 317 #molecular-weight 37642 #checksum 917

Query Match 56.4% Score 53; DB 2; Length 317;
Best Local Similarity 60.0%; Pred. No. 2.85e+01;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 229 VFIRKIEENL 238
Qy 1 VFIRVSNVII 10

```

Search completed: Mon Jun 19 16:17:16 2000  
Total time : 7 secs.

-----

\*\*\*\*\*  
 W A R N I N G  
 (TM)  
 \*\*\*\*\*

Release 3.1A John F. Collins, BioComputing Research Unit,  
 Copyright (c) 1993-1998 University of Edinburgh, U.K.  
 Distribution rights by Oxford Molecular Ltd

Msearch\_pf protein - protein database search, using Smith-Waterman algorithm  
 Run on: Mon Jun 19 16:16:17 2000: Masspar time 3.50 Seconds  
 Iterular output not generated. 112 992 Million cell updates/sec

Title: >US-09-142-524A-13  
 Description: (1-13) from US09142524A.pep  
 Perfect Score: 94  
 Sequence: : VFIRKVSNIITG 13  
 Scoring table: FAM 150  
 Gap 15

Searched: 39857 seqs, 30454973 residues

Post-processing: Minimum Match 0%  
 Listing first 45 summaries

Database: swiss-prot38  
 1:swissprot

Statistics: Mean 27.045; Variance 31.177; scale 0.867

Pred. No. is the number of results predicted by chance to have a  
 score greater than or equal to the score of the result being printed,  
 and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Query Match | Length | ID  | Description | Pred. No. |
|------------|-------------|--------|-----|-------------|-----------|
| 1          | 94          | 100.0  | 1   | SBP-CHYZA   | 7.95e-10  |
| 2          | 61          | 64.9   | 125 | PEL-EXENI   | 1.62e-01  |
| 3          | 59          | 42.8   | 413 | L756-CAEEL  | 4.45e-01  |
| 4          | 57          | 50.6   | 256 | BLAB-BACCE  | 1.19e+00  |
| 5          | 57          | 60.6   | 299 | SUCD-BACSU  | 1.9e+00   |
| 6          | 56          | 59.6   | 257 | BLA2-BACSP  | 1.93e+00  |
| 7          | 56          | 59.6   | 257 | BLA2-BACCE  | 1.93e+00  |
| 8          | 56          | 59.6   | 290 | KPR-LNITU   | 1.93e+00  |
| 9          | 55          | 58.5   | 452 | IES3-LVZVD  | 1.93e+00  |
| 10         | 55          | 58.5   | 550 | YOK7-CAEEL  | 3.11e+00  |
| 11         | 54          | 57.4   | 303 | SUCD-MYCTA  | 3.11e+00  |
| 12         | 53          | 55.4   | 374 | CATB-PSEPU  | 4.98e+00  |
| 13         | 53          | 56.4   | 475 | CYSN-ECOLI  | 7.91e+00  |
| 14         | 53          | 56.4   | 985 | NAH1-YEAST  | 7.91e+00  |
| 15         | 52          | 55.3   | 404 | 9612-LYCES  | 1.25e+01  |
| 16         | 52          | 55.3   | 405 | CPAM-BACSU  | 1.25e+01  |
| 17         | 52          | 55.3   | 405 | THN8-YEAST  | 1.25e+01  |
| 18         | 51          | 54.3   | 261 | SCAL-TRIVA  | 1.95e+01  |
| 19         | 51          | 54.3   | 309 | SCA3-TRIVA  | 1.95e+01  |
| 20         | 51          | 54.3   | 379 | YB8L-YEAST  | 1.95e+01  |
| 21         | 51          | 54.3   | 379 | YF02-YEAST  | 1.95e+01  |
| 22         | 51          | 54.3   | 379 | YF02-YEAST  | 1.95e+01  |
| 23         | 51          | 54.3   | 381 | YNT6-YEAST  | 1.95e+01  |

|    |    |      |      |   |              |                          |          |
|----|----|------|------|---|--------------|--------------------------|----------|
| 24 | 51 | 54.3 | 601  | 1 | YE06_SCHPO   | HYPOTHETICAL 58.8 KD P   | 1.95e-01 |
| 25 | 51 | 54.3 | 720  | 1 | YM52_YEAST   | HYPOTHETICAL 92.1 KD P   | 1.95e-01 |
| 26 | 51 | 54.3 | 785  | 1 | VD05_VACCC   | PROTEIN D5.              | 1.95e-01 |
| 27 | 51 | 54.3 | 785  | 1 | VD05_VACCV   | PROTEIN D5.              | 1.95e-01 |
| 28 | 51 | 54.3 | 785  | 1 | VD05_VAVV    | PROTEIN D5.              | 1.95e-01 |
| 29 | 51 | 54.3 | 810  | 1 | ME8B-BACSU   | NEGATIVE REGULATOR OF    | 1.95e-01 |
| 30 | 51 | 54.3 | 819  | 1 | YU85-INPVP   | HYPOTHETICAL 91.1 KD P   | 1.95e-01 |
| 31 | 51 | 54.3 | 857  | 1 | CLPB-PROTEIN | CLPB PROTEIN (HEAT SHOCK | 1.95e-01 |
| 32 | 51 | 54.3 | 1161 | 1 | DP1A-BORBU   | DNA POLYMERASE III, AC   | 1.95e-01 |
| 33 | 50 | 53.2 | 77   | 1 | RUXA-ARCFU   | PUTATIVE SNRP SM-LIKE    | 3.04e-01 |
| 34 | 50 | 53.2 | 129  | 1 | CYT_CYPCA    | CYSTATIN PRECURSOR (OV   | 3.04e-01 |
| 35 | 50 | 53.2 | 262  | 1 | RS4_CANAL    | 40S RIBOSOMAL PROTEIN    | 3.04e-01 |
| 36 | 50 | 53.2 | 270  | 1 | Y442_MYCGE   | HYPOTHETICAL PROTEIN M   | 3.04e-01 |
| 37 | 50 | 53.2 | 305  | 1 | KIG1-RAI     | PROBABLE G-PROTEIN-GU    | 3.04e-01 |
| 38 | 50 | 53.2 | 587  | 1 | REL-MOUSE    | C-REL PROTO-ONCOGENE P   | 3.04e-01 |
| 39 | 50 | 53.2 | 633  | 1 | NCQG-RHIS    | PROBABLE SULFATE ADENY   | 3.04e-01 |
| 40 | 50 | 53.2 | 639  | 1 | HTPG-PEPA    | HEAT SHOCK PROTEIN HTP   | 3.04e-01 |
| 41 | 50 | 53.2 | 788  | 1 | EXGL-CYCCA   | GLUCAN 1,3-BETA-GLUCOS   | 3.04e-01 |
| 42 | 50 | 53.2 | 908  | 1 | H104_YEAST   | HEAT SHOCK PROTEIN 104   | 3.04e-01 |
| 43 | 50 | 53.2 | 972  | 1 | MIRA_YEAST   | MKNA TRANSPORT REGULAT   | 3.04e-01 |
| 44 | 49 | 52.1 | 310  | 1 | PIP-NEICO    | PROLINE IM:NUDEPTIDASE   | 4.68e-01 |
| 45 | 49 | 52.1 | 981  | 1 | GENE_HAEIN   | SULFAMATE-AMONIA-LIGA    | 4.68e-01 |

## ALIGNMENTS

RESULT 1  
 ID SBP-CHYZA STANDARD; PRI: 374 AA.  
 AC P18532;  
 DT 01-NOV-1990 (Rel. 16, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE SUGI: BASIC PROTEIN PRECURSOR (SBP) (MAJOR ALLERGEN CRY 3.1) (CRY 3.1);  
 OS Cryptomeria japonica (Japanese cedar);  
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC euphyllophytes; Spermatophyta; Coniferopsida; Coniferales;  
 OC Taxodiaceae; Cryptomeria;  
 RN [1]  
 RC SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RP TISSUE=POLLEN;  
 RX MEDLINE: 94193234.  
 RA Sone T., Koniya N., Shimizu K., Kusakabe T., Morikubo K.,  
 RA Kino K.;  
 RT "Cloning and sequencing of cDNA coding for Cry j 1, a major allergen  
 RT of Japanese cedar pollen."  
 RL Biochem. Biophys. Res. Commun 199;519-525(1994).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=POLLEN;  
 RA Namba M., Kuroue M., Torioka K., Fukuda S., Kurimoto M.;  
 RA Submitted (1992-1994) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE OF 22-41.  
 RC TISSUE=POLLEN;  
 RX MEDLINE: 89031257.  
 RA Tanai M., Ando S., Usui M., Kurimoto M., Sakaguchi M., Koike S.,  
 RA Matsuda T.;  
 RT "N-terminal amino acid sequence of a major allergen of Japanese cedar  
 RT pollen (Cry j 1)."  
 RL FEBS Lett. 239;329-332(1988).  
 RN [4]  
 RP CARBOHYDRATES.  
 RX MEDLINE: 95003748.  
 RA Hijikata A., Matsumoto I., Kojima K., Ogawa H.;  
 RT "Antigenicity of the oligosaccharide moiety of the Japanese cedar  
 RT (Cryptomeria japonica) pollen allergen, Cry j 1."  
 RL Int. Arch. Allergy Immunol. 105;198-202(1994).  
 CC -1- PTM: CONTAINS FUCOSE/XYLOSE-CONTAINING N-LINKED OLIGOSACCHARIDES.  
 CC -1- DISEASE: THE MOST COMMON POLLEN ALLERGEN OF JAPANESE CEDAR  
 CC POLLEN. THE MOST COMMON POLLEN ALLERGEN IN JAPAN.  
 CC -1- MISCELLANEOUS: THE SEQUENCE OF CRY j 1 FORM A IS SHOWN HERE. FORM  
 CC B DIFFERS IN SIX POSITIONS.  
 CC -1- SIMILARITY: BELONGS TO THE POLYSACCHARIDE LYASE FAMILY 1.

```

CC      AMB A I/AMB A II/CRY J I SUBFAMILY.
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC
CC      EMBL: D26544; BAA05542.1;
CC      KW Lyase; Signal.
CC      FT SIGNAL 2;
CC      FT CHAIN 22 374;
CC      FT VARIAT 12 12;
CC      FT VARIAT 13 13;
CC      FT VARIAT 202 202;
CC      FT VARIAT 221 221;
CC      FT VARIAT 358 358;
CC      FT VARIAT 361 361;
CC      FT CARBOHYD 148 156;
CC      FT CARBOHYD 191 191;
CC      FT CARBOHYD 243 243;
CC      SEQUENCE 374 AA; 46545 MW; 74AB25950249F56F CRC64;
CC
CC      Query Match 100.0%; Score 94; DB 1; Length 374;
CC      Best Local Similarity 100.0%; Pred. No. 7.99e-10;
CC      Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CC
CC      Db 129 VFIRKRVSNV:IRG 141
CC      QY 1 VFIRKRVSNV:IRG 13
CC
CC      RESULT 2
CC      ID PEL-EMENI STANDARD: PRT: 326 AA.
CC      AC Q00645;
CC      DT 15-DEC-1998 (Rel. 37, Created)
CC      DT 15-DEC-1998 (Rel. 37, Last sequence update)
CC      DT 15-DEC-1998 (Rel. 37, Last annotation update)
CC      DE PECTATE LYASE, PRECURSOR (EC 4.2.2.2).
CC      IN PELA.
CC      CS Emericella nidulans (Aspergillus nidulans).
CC      CC Eukaryota; Fungi; Ascomycota; Plecomycetes; Eurotiiales;
CC      CC Trichomaceae; Emericellaceae.
CC      RN [1]
CC      RP SEQUENCE FROM N.A.
CC      RC STRAIN=PGSC 4;
CC      RX MEDLINE: 95308536.
CC      RA Ho M.C., Whitehead M.P., Cleveland T.E., Dean R.A.;
CC      *Sequence analysis of the Aspergillus nidulans pectate lyase pela
CC      gene and evidence for binding of promoter regions to CREA, a
CC      regulator of carbon catabolite repression.*;
CC      Curr. Genet. 27:142-149(1995).
CC      RL -1- CATALYTIC ACTIVITY: ELIMINATIVE CLEAVAGE OF PECTATE TO GIVE
CC      OLIGOSACCHARIDES WITH 4-DEOXY-ALPHA-D-MANN-4-ENURONOSYL GROUPS
CC      A: THEIR NON-REDUCING ENDS
CC      CC -1- SUBCELLULAR LOCATION: SECRETED (POTENTIAL).
CC      CC -1- SIMILARITY: BELONGS TO THE POLYSACCHARIDE LYASE FAMILY 1.
CC
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC

```

```

DR EMBL: U05592; AAA80568.1;
DR PFAM: PF00544; pec_lyase;
KW Lyase; Signal.
FT SIGNAL 1 15
FT CHAIN 16 326
FT SEQUENCE 326 AA; 34580 MW; F6445A4A6D615D49 CRC64;
CC
CC      Query Match 64.9%; Score 51; DB 1; Length 326;
CC      Best Local Similarity 63.5%; Pred. No. 1.62e-01;
CC      Matches 7; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
CC
CC      Db 110 YIRRVSNV:IRG 120
CC      QY 2 YIRRVSNV:IRG 12
CC
CC      RESULT 3
CC      ID L756_CAEEL STANDARD: PRT: 413 AA.
CC      AC Q11184;
CC      DT 01-NOV-1997 (Rel. 35, Created)
CC      DT 01-NOV-1997 (Rel. 35, Last sequence update)
CC      DT 15-JUL-1998 (Rel. 36, Last annotation update)
CC      DE LEI-756 PROTEIN.
CC      GN LEI-756 OR C05D11.4.
CC      OS Caenorhabditis elegans.
CC      CC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae.
CC      CC Rhabditina; Rhabditidae; Rhabditidae; Pelodermidae; Caenorhabditis.
CC      RN [1]
CC      RP SEQUENCE FROM N.A.
CC      RC STRAIN=BRISTOL NZ;
CC      RA Du 2;
CC      CC -1- SIMILARITY: BELONGS TO THE HEPARIN-BINDING GROWTH FACTORS FAMILY
CC
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC
CC      EMBL: J00048; AAB53825.1;
CC      DR HSSP: P05230; 2AXM.
CC      DR WORMPEP; C05D11.4; CE01134.
CC      DR PROSITE; P00247; HBGF_FGF; FALSE_NEG.
CC      DR PFAM: PF00167; FGF; 1.
CC      SEQUENCE 413 AA; 49116 MW; 78C07D2ACA035B6F CRC64;
CC
CC      Query Match 52.8%; Score 59; DB 1; Length 413;
CC      Best Local Similarity 50.0%; Pred. No. 4.45e-01;
CC      Matches 6; Conservative 5; Mismatches 1; Indels 0; Gaps 0;
CC
CC      Db 353 VFVKRNSDV:IN 364
CC      QY 1 VFVKRNSDV:IN 12
CC
CC      RESULT 4
CC      ID BLAB_BACCE STANDARD: PRT: 256 AA.
CC      AC P14488;
CC      DT 01-JAN-1990 (Rel. 13, Created)
CC      DT 01-JAN-1990 (Rel. 13, Last sequence update)
CC      DT 15-JUL-1999 (Rel. 38, Last annotation update)
CC      DE BETA-LACTAMASE, TYPE II PRECURSOR (EC 3.5.2.6) (PENICILLINASE)
CC      OS Bacillus cereus.
CC      CC Bacteria; Firmicutes; Bacillus/Clostridium group;
CC      CC Bacillus/Staphylococcus group; Bacillus.
CC      RN [1]
CC      RP SEQUENCE FROM N.A.
CC      RC STRAIN=5/6;
CC      RX MEDLINE: 88227879.

```





```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement. (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M15350; AAA22276.1;
DR PIR: A24933; PNB525;
DR HSSP: F04190; 1BMC;
DR PROSITE: PS00743; BETA-LACTAMASE_B_1;
DR PROSITE: PS00744; BETA-LACTAMASE_B_2; 1;
DR PFAM: PF00753; lactamase_B; 1;
KW Hydrolase; Zinc; Antibiotic resistance; Signal;
FT SIGNAL 1 30
FT CHAIN 31 257 BETA-LACTAMASE, TYPE II;
FT METAL 116 116 ZINC 1 (BY SIMILARITY);
FT METAL 118 118 ZINC 1 (BY SIMILARITY);
FT METAL 120 120 ZINC 2 (BY SIMILARITY);
FT METAL 179 179 ZINC 1 (BY SIMILARITY);
FT METAL 198 198 ZINC 2 (BY SIMILARITY);
FT METAL 240 240 ZINC 2 (BY SIMILARITY);
SQ SEQUENCE 257 AA; 28153 MW; 23EEBDC045AE57; CRC64;
Query Match 59.6% Score 56; DB 1; Length 257;
Best Local Similarity 70.0%; Pred. No. 1.93e+00;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
(b 105 FOKRVTDVII 114
1 11111111
2 FIKRVSNVII 11
QY 2 FIKRVSNVII 11
RESULT 7
(b BLA2_NACCE STANDARD; PRT: 257 AA;
AC PQ4190;
FT 20-MAR-1987 (Rel. 34, Greated)
(b 15-FEB-2000 (Rel. 35, Last annotation update)
(b BETA-LACTAMASE, TYPE II PRECURSOR (EC 3.5.2.6) (PENICILLINASE)
(b (CEPHALOSPORINASE);
(b BLK;
(b Bacillus cereus;
(b Bacteria; Firmicutes; Bacillus/Clostridium group;
(b Bacillus/Staphylococcus group; Bacillus;
(b 11)
(b SEQUENCE FROM N.A.
(b STRAIN-569/H;
(b MEDLINE: 86008056;
(b Hussain M., Carling A., Nabenna K.J., Lamphen J.G.;
(b "Cloning and sequencing of the penicillin-binding beta-lactamase II
(b gene of Bacillus cereus 569/H in Escherichia coli.";
(b R. J. Bacteriol. 164:223-229(1985);
(b 12)
(b SEQUENCE OF 31-183; 187-210 AND 214-257;
(b STRAIN-569/H;
(b MEDLINE: 86005446;
(b Ambler R.P., Daniel M., Fleming J., Hermoso J.M., Pang C., Waley S.G.;
(b "The amino acid sequence of the zinc-requiring beta-lactamase II from
(b the bacterium Bacillus cereus 569.";
(b FEBS Lett. 183:207-211(1985);
(b 13)
(b X-RAY CRYSTALLOGRAPHY (3.5 ANGSTROMS);
(b MEDLINE: 88133841;
(b Sutton B.J., Artymix P.J., Cordero-Borboa A.E., Little C.,
(b Phillips D.C., Waley S.G.;
(b "An X-ray-crystallographic study of beta-lactamase II from Bacillus
(b cereus at 0.35-nm resolution.";
(b Biochem. J. 248:181-188(1987);
(b 14)
(b X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS);

```

```

RX MEDLINE: 96067120;
RA Carli A., Pares S., Duee E., Galleni M., Duez C., Frere J.-M.,
RA Dideberg O.;
RT "The 3-D structure of a zinc metallo-beta-lactamase from Bacillus
RT cereus reveals a new type of protein fold.";
RL EMBO J. 14:4914-4921(1995);
RN [5]
RP X-RAY CRYSTALLOGRAPHY (1.85 ANGSTROMS);
RX MEDLINE: 98437525;
RA Carli A., Duee E., Galleni M., Frere J.-M., Dideberg O.;
RT "1.85-A resolution structure of the zinc (II) beta-lactamase from
RT Bacillus cereus.";
RL Acta Crystallogr. D 54:313-323(1998);
RN [6]
RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS);
RX MEDLINE: 98400945;
RA Fabiane S.M., Sohi M.K., Wan I., Payne D.J., Bateson J.H.,
RA Mitchell T., Sutton B.J.;
RT "Crystal structure of the zinc-dependent beta-lactamase from Bacillus
RT cereus at 1.9-A resolution: binuclear active site with features of a
RT mononuclear enzyme.";
RL Biochemistry 37:12404-12411(1998);
RN [7]
RP X-RAY CRYSTALLOGRAPHY (1.7 ANGSTROMS);
RX MEDLINE: 98400945;
RA Fabiane S.M., Sohi M.K., Sutton B.J.;
RT Submitted (SEP-1997) to the PDB data bank;
CC -- FUNCTION: CAN HYDROLYZE CARBAPENEM COMPOUNDS.
CC -- CATALYTIC ACTIVITY: A BETA-LACTAM - H(2)O -> A SUBSTITUTED BETA-
CC AMINO ACID.
CC -- COFACTOR: BINDS TWO ZINC IONS PER MOLECULE. THE ENZYME CAN ALSO
CC FUNCTION WITH ONLY ONE ZINC ION.
CC -- SUBUNIT: MONOMER.
CC -- SIMILARITY: BELONGS TO THE CLASS-B BETA-LACTAMASE FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement. (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M11189; AAA22276.1;
DR PIR: A01006; PNB502;
DR PCB: 1BMC; 28-AUG-96;
DR PCB: 1BME; 07-JUL-97;
DR PCB: 1BV; 23-SEP-99;
DR PCB: 1SC2; 14-OCT-98;
DR PCB: 1SC2; 20-APR-99;
DR PCB: 1SC2; 20-APR-99;
DR PROSITE: PS00743; BETA-LACTAMASE_B_1;
DR PROSITE: PS00744; BETA-LACTAMASE_B_2; 1;
DR PFAM: PF00753; lactamase_B; 1;
KW Hydrolase; Zinc; Antibiotic resistance; Signal; 3D-struct.ref;
FT SIGNAL 1 30
FT CHAIN 31 257 BETA-LACTAMASE, TYPE II;
FT METAL 116 116 ZINC 1, HIGH-AFFINITY;
FT METAL 118 118 ZINC 1, HIGH-AFFINITY;
FT METAL 120 120 ZINC 1, LOW-AFFINITY;
FT METAL 179 179 ZINC 1, HIGH-AFFINITY;
FT METAL 198 198 ZINC 2, LOW-AFFINITY;
FT METAL 240 240 ZINC 2, LOW-AFFINITY;
SQ SEQUENCE 257 AA; 28092 MW; 268EBFB7DDA45431; CRC64;
Query Match 59.6% Score 56; DB 1; Length 257;
Best Local Similarity 70.0%; Pred. No. 1.93e+00;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
(b 105 FOKRVTDVII 114
1 11111111
2 FIKRVSNVII 11
QY 2 FIKRVSNVII 11

```

```

RESULT 8
ID KPRR_NITVU STANDARD: PRT: 290 AA.
AC P37100:
DI 01-OCT-1994 (Rel. 30, Created)
DI 01-OCT-1994 (Rel. 30, Last sequence update)
DI 01-OCT-1994 (Rel. 30, Last annotation update)
DE PHOSPHORIBULOKINASE (EC 2.7.1.19) (PHOSPHOPENTOKINASE) (PRK).
GN CBPP.
OS Nitrobacter vulgaris.
OC Bacteria; Proteobacteria; alpha subdivision: Bradyrhizobium group;
OC Nitrobacteraceae; Nitrobacter.
RN [1]
SE SEQUENCE FROM N.A.
RC STRAIN-T3;
RA Strecker M., Sickinger E., English R.S., Shively J.M., Bock E.;
RL Submitted (XX-1994) to the EMBL/GenBank/DBJ databases.
CC -!- CATALYTIC ACTIVITY: ATP + D-RIBULOSE 5-PHOSPHATE - ADP +
CC D-RIBULOSE 1,5-BISPHOSPHATE.
CC -!- PATHWAY: CALVIN CYCLE.
CC -!- SIMILARITY: BELONGS TO THE PHOSPHORIBULOKINASE FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: L22884; AAA25506.1;
DR HSSP: P12033; IA7J.
DR PR-NTS: PRC0478; PHRIE-KINASE.
DR PROSITE: PS00567; PHOSPHORIBULOKINASE: 1.
EW PFAM: PF00485; PRK: 1.
KW Transferase; Kinase; Calvin cycle; ATP-binding; Photosynthesis.
FT NP_354282.12
SQ SEQUENCE 290 AA: 34908 MW: 952972319425410 CRC64:
Query Match 59.6% Score 55; DB 1; Length 290;
Best Local Similarity 41.7%; Pred. No. 1,93e-00;
Matches 5; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

25 254 PMSRANSVING 265
1 1
2 2 FIKRVSNIH 13

RESULT 9
ID IE61_VZVD STANDARD: PRT: 452 AA.
AC P09269:
DI 01-MAR-1995 (Rel. 10, Created)
DI 01-MAR-1995 (Rel. 10, Last sequence update)
DI 01-APR-1993 (Rel. 25, Last annotation update)
DE TRANSCRIPTIONAL REGULATOR IE63 HOMOLOG.
GN 4.
OS Varicella-zoster virus (strain Dumas) (VZV).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Varicelloviruses.
RN [1]
SE SEQUENCE FROM N.A.
RA MEDLINE: 86306657.
RX Davison A.J., Scott J.E.;
RT "The complete DNA sequence of varicella-zoster virus."
RL J. Gen. Virol. 67:1759-1816(1986).
CC -!- SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV-1 UL54,
CC HSV-2 UL54, EBV-1 5, VZV 4, EBV BMLF1, HCMV UL65, AND HSV-1 57.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: X04370; CAA27887.1;
DR PIR: D27212; WZBE4.
KW Transcription regulation.
SQ SEQUENCE 452 AA: 51543 MW: 429265471E380B4 CRC64:
Query Match 58.5% Score 55; DB 1; Length 452;
Best Local Similarity 38.5%; Pred. No. 3.11e-00;
Matches 5; Conservative 6; Mismatches 2; Indels 0; Gaps 0;

DB 370 VMIARIANIVVRG 382
1 1
2 1 VFIKRVSNV 13

RESULT 10
ID YQK7_CAEEL STANDARD: PRT: 550 AA.
AC Q09289;
DI 01-NOV-1997 (Rel. 35, Created)
DI 01-NOV-1997 (Rel. 35, Last sequence update)
DI 01-NOV-1997 (Rel. 35, Last annotation update)
DE HYPOTHETICAL 59.4 KD PROTEIN C56G2.7 IN CHROMOSOME 111.
GN C56G2.7.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae.
OC Rhabditia; Rhabditidae; Rhabditidae; Peleodidae; Caenorhabditis
RN [1]
SE SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Connell M.;
RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: TO HUMAN I10 KD CELL MEMBRANE GLYCOPROTEIN
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U23177; AAA64334.1;
DR WORMPEP: C56G2.7; CE01876.
KW Hypothetical protein.
SQ SEQUENCE 550 AA: 59367 MW: 87847A58EF0AEE8 CRC64:
Query Match 58.5% Score 55; DB 1; Length 550;
Best Local Similarity 50.0%; Pred. No. 3.11e-00;
Matches 5; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

DB 224 VFIKOSNDMLH 235
1 1
2 1 VFIKRVSNV 13

RESULT 11
ID SUCD_MYCTC STANDARD: PRT: 303 AA.
AC P71558;
DI 01-NOV-1997 (Rel. 35, Created)
DI 01-NOV-1997 (Rel. 35, Last sequence update)
DI 15-FEB-2000 (Rel. 39, Last annotation update)
DE SUCCINYL-CCA SYNTHETASE ALP1 CHAIN (EC 6.2.1.5) (SCS-ALPHA).
GN SUCD OR RV0952 OR MTCY10D7.22C.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium
RN [1]
SE SEQUENCE FROM N.A.
RC STRAIN-H37RV;
RX MEDLINE: 98295987.
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,

```

RA Gordon S.V., Eigelmeier K., Gas S., Barry C.E. III, Tekala F.,  
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
 RA Davies K., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,  
 RA Hornsby T., Jacobs K., Krogh A., McLean J., Moule S., Murphy L.,  
 RA Oliver S., Osborne K., Quail M.A., Rajandream M.A., Rogers J.,  
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R., Sulston J.E.,  
 RA Taylor K., Whitehead S., Barrett B.G.,  
 RA "Deciphering the biology of Mycobacterium tuberculosis from the  
 RA complete genome sequence."  
 RL Nature 393:537-544(1998).  
 CC -!- CATALYTIC ACTIVITY: SUCCINATE + COA + ATP -> SUCCINYL-COA + ADP +  
 CC ORTHOPHOSPHATE.  
 CC -!- SUBUNIT: Tetramer of two alpha and two beta chains  
 CC (BY SIMILARITY)  
 CC -!- SIMILARITY: TO OTHER ALPHA SUBUNITS OF SUCCINYL-COA SYNTHETASE.  
 CC OF MALATE-COA LIGASE AND TO ATP CITRATE-LIASE.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).)  
 CC -----  
 CC EMBL: Z79700; CAB01959.1;  
 CC HSSP: PC7459; ISCU.  
 CC PROSITE: PS01216; SUCCINYL\_COA\_LIG\_1; 1.  
 CC PROSITE: PS03499; SUCCINYL\_COA\_LIG\_2; 1.  
 CC PFAM: PF00549; ligase-coa; 1.  
 CC TUBERCULOSIS: K00952; .  
 CC KW Ligase; Tricarboxylic acid cycle; ATP-binding; Phosphorylation.  
 CC LIASE; Tricarboxylic acid cycle; PHOSPHORYLATED IN THE COURSE OF  
 CC CATALYSIS (BY SIMILARITY).  
 CC ACT\_SITE 259 259  
 CC SEQUENCE 303 AA; 31229 MW; 57769CE8DCB48 CRC64;  
 CC -----  
 CC Query Match 57.4%; Score 54; DB 1; Length 303;  
 CC Best Local Similarity 38.5%; Pred. No. 4.96e+00;  
 CC Matches 5; Conservative 5; Mismatches 2; Indels 0; Gaps 0;  
 CC -----  
 CC 6 IFLSKDKVKVVG 18  
 CC 1 VFIKRVSNVTH 13  
 CC -----  
 CC RESULT 12  
 CC ID CATB\_PSEP0 STANDARD; PRT: 374 AA;  
 CC AC P08310;  
 CC DT 01-AUG-1988 (Rel. 08, Created);  
 CC DT 01-MAY-1991 (Rel. 18, Last sequence update);  
 CC DT 15-JUL-1995 (Rel. 38, Last annotation update);  
 CC DE MUONATE CYCLOISOMERASE I (EC 5.5.1.1) (CIS-CIS-MUONATE LACTONIZING  
 CC ENZYME I) (MFE);  
 CC GN CATB;  
 CC OS Pseudomonas putida.  
 CC OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonas group;  
 CC CC Pseudomonas.  
 CC (1)  
 CC SEQUENCE FROM N.A.  
 CC STRAIN-PRS2015;  
 CC MEDLINE: 8727391.  
 CC RA Aldrich T.L., Frantz B., Gill J.F., Kibane J.J., Chakrabarty A.M.;  
 CC "Cloning and complete nucleotide sequence determination of the catB  
 CC gene encoding cis,cis-muonate lactonizing enzyme."  
 CC J. Bacteriol. 170:1297-1304(1988).  
 CC (2)  
 CC SEQUENCE FROM N.A.  
 CC MEDLINE: 88139192.  
 CC RA Aldrich T.L., Chakrabarty A.M.;  
 CC "Transcriptional regulation, nucleotide sequence, and localization of  
 CC the promoter of the catB operon in Pseudomonas putida."  
 CC J. Bacteriol. 170:1297-1304(1988).  
 CC (3)

RP X-RAY CRYSTALLOGRAPHY (3.0 ANGSTROMS).  
 RX MEDLINE: 87283908.  
 RA Goldman A., Ollis D.L., Steitz T.A.;  
 RT "Crystal structure of muonate lactonizing enzyme at 3-A resolution."  
 RL J. Mol. Biol. 194:143-153(1987).  
 CC (4)  
 RN X-RAY CRYSTALLOGRAPHY (1.85 ANGSTROMS).  
 RC STRAIN-PRS2000;  
 RA Hedlin S., Kahn P.C., Guha B., Malloes D.J., Goldman A.;  
 RT "The refined X-ray structure of muonate lactonizing enzyme from  
 RT Pseudomonas putida PRS2000 at 1.85-A resolution."  
 RL J. Mol. Biol. 254:918-941(1995).  
 CC (5)  
 RN SIMILARITY TO MR  
 RP MEDLINE: 91015392.  
 RA Neidhart D.J., Kenyon G.L., Gerit J.A., Petsko G.A.;  
 RT "Mandelate racemase and muonate lactonizing enzyme are  
 RT mechanistically distinct and structurally homologous."  
 RL Nature 347:692-694(1990).  
 CC -!- FUNCTION: CATALYZES A SYN CYCLOISOMERIZATION.  
 CC -!- CATALYTIC ACTIVITY: 2,5-DIHYDRO-5-OXOFURAN-2-ACETATE -> CIS-CIS-  
 CC HEXADIENEDIOLATE.  
 CC -!- COFACTOR: REQUIRES MANGANESE.  
 CC -!- PATHWAY: SECOND STEP IN THE DEGRADATION OF CATECHOL TO SUCCINATE  
 CC AND ACETYL-COA IN THE BETA-KETOCARBOXYLATE PATHWAY.  
 CC -!- SUBUNIT: HOMODIMER.  
 CC -!- SIMILARITY: BELONGS TO THE MANDELATE RACEMASE / MUONATE  
 CC LACTONIZING ENZYME FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).)  
 CC -----  
 CC EMBL: M16560; AAA25765.1;  
 CC EMBL: M16336; AAA25765.1;  
 CC PIR: A27316; A27316.  
 CC PIR: A28530; A28530.  
 CC PDB: 1MUC; 11-JUL-95.  
 CC PDB: 2MUC; 04-NOV-98.  
 CC PDB: 3MUC; 04-NOV-98.  
 CC PROSITE: PS00908; MR\_MLE\_1; 1.  
 CC PROSITE: PS00909; MR\_MLE\_2; 1.  
 CC PFAM: PF01188; MR\_MLE; 1.  
 CC KW Aromatic hydrocarbons catabolism; Isomerase; Manganese; Cysteine;  
 CC ACT\_SITE 170 170  
 CC METAL 199 199  
 CC METAL 225 225  
 CC METAL 250 250  
 CC CONFLICT 32 32 T -> S (IN REF. 2).  
 CC CONFLICT 64 64 S -> I (IN REF. 2).  
 CC CONFLICT 243 243 MISSING (IN REF. 2).  
 CC SEQUENCE 374 AA; 40378 MW; A7767A865EACD8C CRC64;  
 CC -----  
 CC Query Match 56.4%; Score 53; DB 1; Length 374;  
 CC Best Local Similarity 41.7%; Pred. No. 7.91e+00;  
 CC Matches 5; Conservative 4; Mismatches 3; Indels 0; Gaps 0;  
 CC -----  
 CC 3 VLIETETAIIVH 14  
 CC 1 VFIKRVSNVTH 12  
 CC -----  
 CC RESULT 13  
 CC ID CYSN\_EC01 STANDARD; PRT: 475 AA;  
 CC AC P23845;  
 CC DT 01-NOV-1991 (Rel. 20, Created)  
 CC DT 01-NOV-1991 (Rel. 20, Last sequence update)

15-DEC-1998 (Rel. 37, Last annotation update)  
 DE SULFATE ADENYLATE TRANSFERASE SUBUNIT 1 (EC 2.7.7.4) (ATP-  
 DE SULFURYLASE).  
 GN CYSN.  
 OS Escherichia coli.  
 OC Bacteria: Proteobacteria: gamma subdivision: Enterobacteriaceae:  
 OC Escherichia.  
 RN [1]  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-5.  
 RC STRAIN-K12:  
 RX MEDLINE: 92426517.  
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
 RA Riley M., Collado-Vides J., Glasner F.D., Rode C.K., Mayhew G.F.,  
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
 RA Mau B., Shao Y.,  
 RA "The complete genome sequence of Escherichia coli K-12".  
 RL Science 277:1453-1474 (1997).  
 RN [3]  
 RP CHARACTERIZATION.  
 RX MEDLINE: 89115389.  
 RA Levy T.S., Taylor J.C., Markham G.D.,  
 RA "The sulfate activation locus of Escherichia coli K12: cloning,  
 RI genetic, and enzymatic characterization".  
 RL J. Biol. Chem. 263:2409-2416 (1988).  
 CC -!- FUNCTION: MAY BE THE GTPASE, REGULATING ATP SULFURYLASE ACTIVITY.  
 CC -!- CATALYTIC ACTIVITY: ATP + SULFATE -> PYROPHOSPHATE +  
 CC ADENYLYLSULFATE.  
 CC -!- ENZYME REGULATION: GTPASE ACTIVITY IS COUPLED TO STIMULATION OF  
 CC THE RATE OF APS FORMATION.  
 CC -!- PATHWAY: FIRST STEP IN THE SULFATE ACTIVATION PATHWAY, THIS  
 CC REACTION OCCURS EARLY IN THE REDUCTIVE BRANCH OF THE CYSTEINE  
 CC BIOSYNTHETIC PATHWAY.  
 CC -!- SUBUNIT: HETERODIMER COMPOSED OF CYSD, THE SMALLER SUBUNIT, AND  
 CC CYSN.  
 CC -!- SIMILARITY: BELONGS TO THE GTP-BINDING ELONGATION FACTOR FAMILY.  
 CC CYSN/NDQD SUBFAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL: M74596; AAA23646.1;  
 DR EMBL: U29579; AAA69261.1;  
 DR EMBL: A5000358; AAC75793.1;  
 DR PIR: JN0327; JN0327.  
 DR HSP: P02990; LETU.  
 DR EC02DBASE; C058.3; 6TH EDITION.  
 DR ECOGENE; EG:0194; CYSN.  
 DR PRINTS; PR00315; ELONGATNFT.  
 DR PROSITE; PS00301; EFATOR\_GTP. 1.  
 DR PFAM; PF00009; GTP\_EFTU. 1.  
 KW Cysteine biosynthesis: Transferase: Nucleotidyltransferase:  
 KW GTP-binding.  
 FT NP\_BIND 34 41 GTP (BY SIMILARITY).  
 FT NP\_BIND 113 117 GTP (BY SIMILARITY).  
 FT NP\_BIND 169 171 GTP (BY SIMILARITY).  
 CC SEQUENCE 475 AA: 52558 MW: 915983649732A15 CRC64:

Query Match 56.4% Score 53: DB 1: Length 475;  
 Best Local Similarity 54.5% Pred. No. 7.91e+00;  
 Matches 6: Conservative 3: Mismatches 2: Indels 0: Gaps 0:

DB 419 IFIDRLSNVTY 429  
 QY 1 VFIRKVSNTY 11  
 RESULT 14  
 ID NAHL YEAST STANDARD: PRT: 985 AA.  
 AC O99271.  
 DI 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 15-FEB-2000 (Rel. 39, Last annotation update)  
 DE NA(+)/H(+), ANTI-PORTER.  
 GN NAH1 OR YLR138W OR L3149 OR L9606.4.  
 OS Saccharomyces cerevisiae "Baker's yeast".  
 OC Eukaryota: Fungi: Ascomycota: Saccharomycetes: Saccharomycetales:  
 OC Saccharomycetaceae: Saccha. myces.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-S288C / AB972:  
 RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Du Z.,  
 RA Favello A., Fulton L., Gattuso D., Greco T., Kirsten J., Kucaba T.,  
 RA Hallsworth K., Hawkins J., Hillier L., Jier M., Johnson D.,  
 RA Johnston L., Langston T., Latvala P., Le T., Mardis E., Menezes S.,  
 RA Miller N., Nhan M., Pauley A., Peluso D., Rifkin L., Riles L.,  
 RA Taich A., Trevisan E., Vignati D., Wilcox L., Woldman P., Vaudin M.,  
 RA Wilson R., Waterston R.,  
 RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-S288C:  
 RA Delius H., Hesling U.,  
 RL Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE OF 145-935 FROM N.A.  
 RA Rieger M., Mueller-Auer S., Brueckner M.,  
 RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.  
 CC -!- FUNCTION: SODIUM EXPORT FROM CELL. TAKES UP EXTERNAL PHOSPHON IN  
 CC EXCHANGE FOR INTERNAL SODIUM IONS (BY SIMILARITY).  
 CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).  
 CC -!- SIMILARITY: BELONGS TO THE FUNGAL NA(+)/H(+), EXCHANGER FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL: U53841; AA82392.1;  
 DR EMBL: X91258; CA62653.1;  
 DR EMBL: Z73310; CA62709.1;  
 DR EMBL: Z73311; CA62711.1;  
 DR SCS; L0004113; NAH1.  
 KW Transmembrane; Glycoprotein; Sodium transport; Transport; Symport.  
 FT TRANSMEM 13 33 POTENTIAL.  
 FT TRANSMEM 37 57 POTENTIAL.  
 FT TRANSMEM 71 91 POTENTIAL.  
 FT TRANSMEM 106 126 POTENTIAL.  
 FT TRANSMEM 129 149 POTENTIAL.  
 FT TRANSMEM 177 197 POTENTIAL.  
 FT TRANSMEM 204 224 POTENTIAL.  
 FT TRANSMEM 245 265 POTENTIAL.  
 FT TRANSMEM 295 315 POTENTIAL.  
 FT TRANSMEM 320 340 POTENTIAL.  
 FT TRANSMEM 362 382 POTENTIAL.  
 FT TRANSMEM 411 431 POTENTIAL.  
 CC SEQUENCE 985 AA: 109369 MW: AD0B86D483DCDD89 CRC54:

Query Match 56.4% Score 53: DB 1: Length 985;  
 Best Local Similarity 36.4% Pred. No. 7.91e+00;  
 Matches 4: Conservative 6: Mismatches 1: Indels 0: Gaps 0:

```

DB 338 IFERRIPAVWI 348
QY 1 VF:KRVSNV::11

RESULT 15
ID 9612 LYCES STANDARD: PFI: 404 AA.
AC P24396:
DI 01-MAR-1992 (Rel. 21: Created)
DI 01-MAR-1992 (Rel. 21: Last sequence update)
DI 15-JUL-1999 (Rel. 38: Last annotation update)
DE STYLE DEVELOPMENT-SPECIFIC PROTEIN 9612 PRECURSOR.
GN 9612.
OS Lycopersicon esculentum (Tomato).
OC Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta:
OC euphyllophytes: Spermatophyta: Magnoliophyta: eudicotyledons:
OC core eudicots: Asteridae: euasterids I: Solanales: Solanaceae:
OC Solanum.
RN [1]
R1 SEQUENCE FROM N.A.
R2 STRAIN: CV_VF36: TISSUE: PISTIL.
R3 MEDLINE: 9117185.
R4 Ruediger K.A., Smith A.G., Gasser C.S.
R5 -Regulation of a stylar transmitting tissue-specific gene in
R6 wild-type and transgenic tomato and tobacco.*
R7 Mol. Gen. Genet. 224:183-192(1990).
R8 -1- FUNCTION: MAY HAVE A ROLE IN THE DEVELOPMENT OF THE TRANSMITTING
R9 TISSUE OF THE STYLE AND/OR IN THE EVENTS RELATED TO POLLINATION
R10 SUCH AS SOME ASPECT IN THE FACILITATION OF COMPATIBLE POLLEN TUBE
R11 GROWTH.
R12 -1- SUBCELLULAR LOCATION: SECRETED.
R13 -1- TISSUE SPECIFICITY: PREDOMINANTLY FOUND IN THE PISTIL WHERE IT IS
R14 FOUND IN THE OUTER FIVE LAYERS OF THE STRANDS OF TRANSMITTING
R15 TISSUE WITHIN THE UPPER TWO-THIRDS OF THE STYLE. FOUND AT MUCH
R16 LOWER LEVELS IN THE ANTHEXIS AND VEGETATIVE ORGANS.
R17 -1- DEVELOPMENTAL STAGE: MAXIMUM LEVELS ARE FOUND DURING ANTHESIS.
R18 -1- SIMILARITY: 54% IDENTICAL TO TOMATO PROTEINS P59 (AC P15722)
R19 AND P56 (AC P15721).
R20
R21 This SWISS-PROT entry is copyright. It is produced through a collaboration
R22 between the Swiss Institute Of Bioinformatics and the EMBL Outstation -
R23 the European Bioinformatics Institute. There are no restrictions on its
R24 use by non-profit institutions as long as its content is in no way
R25 modified and this statement is not removed. Usage by and for commercial
R26 entities requires a license agreement (See http://www.isb-sib.ch/announce/
R27 or send an email to license@isb-sib.ch).
R28
R29 EMBL: X55193; CAA3579.1.1.
R30 F1: S12209; S12209.
R31 PFAM: PF00544; pec_lyuse1.1.
R32 PRINTS: PR03907; AMBALEKGEN.
R33
R34 SIGTAL: 1 20 POTENTIAL.
R35 CHAIN 21 404 STYLE DEVELOPMENT-SPECIFIC PROTEIN 9612.
R36 CARBOHYD 37 37 POTENTIAL.
R37 CARBOHYD 191 191 POTENTIAL.
R38
R39 SEQUENCE 404 AA: 44298 MW: 32660598:2808675 CRO64:
R40
R41 Query Match 55.38: Score 52: DB 1: Length 404;
R42 Best Local Similarity 63.68: Pred. No. 1:25e-01;
R43 Matches 7: Conservative 2: Mismatches 2: Indels 0: Gaps 0:
R44
R45 DB 162 IHHTSNII:HG 172
R46 : ||| |||
R47 QY 3 :KRVSNV::HG 13

```

Search completed: Mon Jun 19 16:16:23 2000  
 Net time : 6 secs.

\*\*\*\*\*  
[WATERBURY] (TM)  
\*\*\*\*\*

Release 3.1A John F. Collins, Bioinformatics Research Unit,  
Copyright (c) 1993-1998 University of Edinburgh, G.K.  
Distribution rights by Oxford Molecular Ltd

WATERBURY protein - protein database search, using Smith-Waterman algorithm  
Run on: Mon Jun 19 16:16:40 2000; Maspar time 8.57 Seconds  
Tubular Output not generated. 105,119 Million cell updates/sec

Title: US-09-142-524A-13  
Description: (1-13) from US09142524A.pep  
Perfect Score: 94  
Sequence: 1 VF:KPVSNVVIHG 13

Scoring table: PAM 150  
Gap 15

Searched: 225878 seqs, 69334122 residues  
Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: spireb112  
1:sp-archaea 2:sp-bacteria 3:sp-fungi 4:sp-human  
5:sp-invertebrate 6:sp-mammal 7:sp-mhc 8:sp-organelle  
9:sp-phage 10:sp-plant 11:sp-protozoa 12:sp-unclassified  
13:sp-vertebrate 14:sp-virus

Statistics: Mean 26.111; Variance 31.885; scale 0.919

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

| SUMMARIES  |       |              |         | SUMMARIES               |           |  |
|------------|-------|--------------|---------|-------------------------|-----------|--|
| Result No. | Score | Match Length | DB ID   | Description             | Pred. No. |  |
| 1          | 84    | 68.1         | 598 14  | ORF MSV077 HYPOTHETICAL | 9.02e-02  |  |
| 2          | 63    | 67.0         | 360 2   | HYPOTHETICAL 41.5 KD P  | 1.50e-01  |  |
| 3          | 61    | 64.9         | 367 10  | POLLEN MAJOR ALLERGEN   | 4.07e-01  |  |
| 4          | 61    | 64.9         | 375 10  | CHAGI PRECURSOR         | 4.07e-01  |  |
| 5          | 59    | 62.8         | 390 10  | FL1P1.22 PROTEIN        | 1.08e+00  |  |
| 6          | 58    | 61.7         | 783 2   | BETA-N-ACETYLGLUCOSAMIN | 1.76e+00  |  |
| 7          | 56    | 59.6         | 408 2   | GLYCOSYLTRANSFERASE GT  | 4.52e+00  |  |
| 8          | 56    | 59.6         | 512 2   | POTASSIUM FLIPASE       | 4.52e+00  |  |
| 9          | 56    | 59.6         | 361 5   | K07A12.2 PROTEIN        | 4.52e+00  |  |
| 10         | 55    | 59.5         | 317 2   | PROLYL AMINOPEPTIDASE   | 7.20e+00  |  |
| 11         | 55    | 58.5         | 1894 10 | TIF15.3 PROTEIN (EC 2.  | 7.20e+00  |  |
| 12         | 54    | 57.4         | 451 10  | MAT1                    | 1.14e+01  |  |
| 13         | 53    | 56.4         | 75 2    | VIRULENCE ASSOCIATED P  | 1.79e+01  |  |
| 14         | 53    | 56.4         | 190 2   | TRANSCRIPTIONAL REGULA  | 1.79e+01  |  |
| 15         | 53    | 56.4         | 317 2   | THYMIDYLATE SYNTHASE C  | 1.79e+01  |  |
| 16         | 53    | 56.4         | 327 5   | CODED FOR BY C. ELEGAN  | 1.79e+01  |  |
| 17         | 53    | 56.4         | 357 14  | HYPOTHETICAL 41.2 KD P  | 1.79e+01  |  |
| 18         | 53    | 56.4         | 405 10  | PTTA                    | 1.79e+01  |  |
| 19         | 53    | 56.4         | 411 10  | PLASMOD GR0497C, COMPL  | 1.79e+01  |  |
| 20         | 53    | 56.4         | 674 5   | C27A7.3 PROTEIN         | 1.79e+01  |  |

|    |    |      |      |    |         |                        |          |
|----|----|------|------|----|---------|------------------------|----------|
| 21 | 53 | 56.4 | 835  | 2  | Q56012  | CELL SURFACE-ASSOCIATE | 1.79e-01 |
| 22 | 53 | 56.4 | 871  | 3  | Q74833  | DNA LIGASE PROTEIN     | 1.79e-01 |
| 23 | 52 | 55.3 | 192  | 14 | Q81434  | ENVELOPE PROTEIN E1 (F | 2.79e-01 |
| 24 | 52 | 55.3 | 394  | 10 | Q65457  | PECTATE LYASE LIKE PRO | 2.79e-01 |
| 25 | 52 | 55.3 | 394  | 10 | Q65456  | PECTATE LYASE LIKE PRO | 2.79e-01 |
| 26 | 52 | 55.3 | 405  | 10 | Q24216  | PECTATE LYASE          | 2.79e-01 |
| 27 | 52 | 55.3 | 427  | 10 | Q19815  | F17A14.4 PROTEIN       | 2.79e-01 |
| 28 | 51 | 54.3 | 111  | 2  | C 2882  | HYPOTHETICAL 12.2 KD P | 4.33e-01 |
| 29 | 51 | 54.3 | 186  | 2  | C 1109  | HYPOTHETICAL 20.9 KD P | 4.33e-01 |
| 30 | 51 | 54.3 | 238  | 5  | C 1411  | COSMID PDB1            | 4.33e-01 |
| 31 | 51 | 54.3 | 254  | 1  | C 1799  | 254AA LONG HYPOTHETICA | 4.33e-01 |
| 32 | 51 | 54.3 | 309  | 5  | C 15568 | ADHESIN PROTEIN AP33-1 | 4.33e-01 |
| 33 | 51 | 54.3 | 309  | 5  | C 15569 | ADHESIN PROTEIN AP33-3 | 4.33e-01 |
| 34 | 51 | 54.3 | 347  | 5  | Q97246  | PCG045W PROTEIN        | 4.33e-01 |
| 35 | 51 | 54.3 | 365  | 5  | Q77241  | HEXAMERIN (FRAGMENT)   | 4.33e-01 |
| 36 | 51 | 54.3 | 417  | 5  | Q18095  | F17H7.10 PROTEIN       | 4.33e-01 |
| 37 | 51 | 54.3 | 581  | 5  | C 22977 | COSMID F45F2           | 4.33e-01 |
| 38 | 51 | 54.3 | 785  | 14 | Q93119  | PUTATIVE 90.4K PROTEIN | 4.33e-01 |
| 39 | 51 | 54.3 | 1006 | 2  | C 67588 | ATPASE SUBUNIT OF ATP  | 4.33e-01 |
| 40 | 50 | 53.2 | 157  | 2  | C 54168 | PEPTIDYLPROLYL ISOMERA | 5.68e-01 |
| 41 | 50 | 53.2 | 225  | 14 | Q24989  | ENVELOPE POLYPROTEIN   | 5.68e-01 |
| 42 | 50 | 53.2 | 260  | 1  | C 1009  | CONSERVED HYPOTHETICAL | 5.68e-01 |
| 43 | 50 | 53.2 | 286  | 5  | C 1447  | H1ZC20.4 PROTEIN       | 5.68e-01 |
| 44 | 50 | 53.2 | 439  | 5  | C 6197  | REVERSE TRANSCRIPTASE  | 5.68e-01 |
| 45 | 50 | 53.2 | 450  | 10 | C 10319 | PECTATE LYASE HOMOLOG  | 5.68e-01 |

ALIGNMENTS

RESULT ID Q9YW15 PRELIMINARY: PRT: 598 AA.  
AC Q9YW15:  
DT 01-MAY-1999 (TrEMBLrel. 10, Created)  
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)  
DT 01-MAY-1999 (TrEMBLrel. 10, Last annotation update)  
DE ORF MSV077 HYPOTHETICAL PRO1 N.  
GN MSV077.  
OS Melanoplus sanguinipes entomopoxvirus.  
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Entomopoxvirinae;  
OC Entomopoxvirus B.  
[1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-TUCSON.  
RX MEDLINE: 9910262.  
RA AFONSO C.L., TULMAN E.R., LU Z., OMA E., KUTISH G.F., ROCK D.L.  
RI The genome of Melanoplus sanguinipes entomopoxvirus.  
RL J. Virol. 73:533-552(1999).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-TUCSON.  
RA AFONSO C.L., TULMAN E.R., LU Z., OMA E., KUTISH G.F., ROCK D.L.  
RI Submitted (MAY-1998) to the EMBL/GenBank/CCRB databases.  
RL EMBL: AF083866; AA: 576321.  
SQ SEQUENCE 598 AA: 73312 MW: 14428F93 CRC32:  
Query Match 68.1%; Score 64; DB 14; Length 598.  
Best Local Similarity 61.5%; Pred. No. 9.02e-02;  
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

DB 105 VIKNVSTIIIDG 117  
QY 1 VF:KPVSNVVIHG 13  
RESULT 2  
ID Q9YGL9 PRELIMINARY: PRT: 360 AA.  
AC Q9YGL9:  
DT 01-MAY-1999 (TrEMBLrel. 10, Created)  
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)  
DT 01-MAY-1999 (TrEMBLrel. 10, Last annotation update)  
DE HYPOTHETICAL 41.5 KD PROTEIN  
OS Leptospira borgpetersenii  
OC Bacteria; Spirochaetales; Leptospiraceae; Leptospira.

```

FN 121
RP SEQUENCE FROM N.A.
RC STRAIN=L171.
RA KAJAMBAHETI T., RULACH D.M., RAJAKUMAR K., ADLER B.:
RT "Genetic Organization of the Lipopolysaccharide O-antigen Biosynthetic
RI Locus of Leptospira borgpetersenii Serovar Hardjohovis.",
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
CR EMBL: AF078135; AAC12950.1; -.
KW Hypothetical protein.
SQ SEQUENCE 360 AA: 41545 MW: 52201EE3 CRC32:
      67.0%   Score 63; DB 2; Length 360;
      Best Local Similarity 56.7%; Pred. NO. 4.07e-01;
      Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 128 FIKRVGVILSG 139
      1111111111
QY 2 FIKRVSNVILHG 13

RESULT 3 PRELIMINARY: PRT: 367 AA.
ID Q92N07
AC Q92N07
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DI 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE POLLEN MAJOR ALLERGEN 1-2.
OS Juniperus ashei (Dark white cedar).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Coniferopsida; Coniferales;
OC Taxodiaceae; Juniperus.
RN [1]
RP SEQUENCE FROM N.A.
RA MIDORO-HORIUTI T.M., GOLDBLUM R.M., KUROSKY A., WOOD T.G.,
RA BROOKS E.G.:
RI "Molecular cloning of mountain cedar (Juniperus ashei) pollen major
RI allergen, Jun a 1.",
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
CR EMBL: AF106663; AAD03609.1; -.
CR EMBL: AF106662; AAD03608.1; -.
CR MENDL: 36544; Juncus:1088;36544.
CR MENDL: 36545; Juncus:1088;36545.
SQ SEQUENCE 367 AA: 19824 MW: 40209610 C 132:
      64.9%   Score 61; DB 10; Length 367;
      Best Local Similarity 46.2%; Pred. NO. 4.07e-01;
      Matches 6; Conservative 7; Mismatches 3; Indels 0; Gaps 0;

Db 129 FIKRVSNVILHG 141
      1111111111
QY 1 FIKRVSNVILHG 13

RESULT 4 PRELIMINARY: PRT: 375 AA.
ID Q96385
AC Q96385
DT 01-FEB-1997 (TrEMBLrel. 02, Created)
DI 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE CHAO1 PRECURSOR.
OS Chamaecyparis obtusa.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Coniferopsida; Coniferales;
OC Taxodiaceae; Chamaecyparis.
RN [1]
RP SEQUENCE FROM N.A.
RA SUZUKI M., KOMIYAMA N., ITOH M., ITOH H., SONE T., KUNO K., TAKAGI I.,
RA OHITA N.:
RT "Purification, characterization and molecular cloning of Chao 1, a
RT major allergen of Chamaecyparis obtusa (Japanese cypress) pollen.",
RL Mol. Immunol. 33:451-460(1996).

```

```

DR EMBL: D45404; BAAC8246.1; -.
DR MENDL: 7625; Chaob:1088;7626.
DR PFAM: PF00544; pec_lyase; 1.
DR PRINTS: PRO0807; AMBALLERGEN.
KW Signal.
FT SIGNAL 1 21 POTENTIAL.
FT CHAIN 22 375 CHAO1.
SQ SEQUENCE 375 AA: 40258 MW: A0981492 CRC32:
      64.9%   Score 61; DB 10; Length 375;
      Best Local Similarity 53.8%; Pred. NO. 4.07e-01;
      Matches 7; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Db 129 LFMRTVSHVILHG 141
      1111111111
QY 1 VF:IKRVSNVILHG 13

RESULT 5 PRELIMINARY: PRT: 390 AA.
ID Q65388
AC Q65388
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DI 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE F12F.22 PROTEIN.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Arabidopsia; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Magnoliopsida; eudicotyledons;
OC core eudicots; Rosidae; eurosids 1; Brassicales; Brassicaceae;
OC Arabidopsis.
RN [1]
RP SEQUENCE FROM N.A.
RA VYSOTSKAYA V.S., OSBORNE B.L., SCHWARTZ J.R., TORIUMI M., YU S.,
RA KWAN A., GIL O., LIU S., BUEHLER E., GUNWAY A.B., GUNWAY A.R.,
RA DEWAR K., FENG J., KIM C., KURTZ D., PALM C.J., LI Y., SHINN P.,
RA SUN H., DAVIS R.W., ECKER J.R., FEDEMSPIEL N.A., THEOLOGIS A.:
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
CR EMBL: AC002131; AAC17625.1; -.
CR MENDL: 29024; Arabid:1088;29024.
CR PFAM: PF00544; pec_lyase; 1.
CR PRINTS: PRO0807; AMBALLERGEN.
SQ SEQUENCE 390 AA: 43354 MW: E0322F54 CRC32:
      64.9%   Score 61; DB 10; Length 390;
      Best Local Similarity 72.7%; Pred. NO. 4.07e-01;
      Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 144 LFMRTVSHVILHG 154
      1111111111
QY 3 IKRVSNVILHG 13

RESULT 6 PRELIMINARY: PRT: 783 AA.
ID Q92H39
AC Q92H39
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DI 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE BETA-N-ACETYLGLUCOSAMINIDASE.
OS Pseudoalteromonas sp. S9.
OC Bacteria; Proteobacteria; gamma subdivision; Alteromonadaceae;
OC Pseudoalteromonas.
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=S9.
RA TECHKARNANARUK S., GOODMAN A.E.:
RT "Cloning, sequence analysis and characterisation of genes involved in
RT chitin degradation of a marine bacterium, Pseudoalteromonas sp. strain
RT S9.",
RL Microbiology 130:0(1999).

```

501 FIKRVINIVFN 51:  
 |||| :  
 2 FIKRVNVIIE 12

Best Local Similarity 41.7%; Pred. No. 7.20e+00;  
Matches 5; Conservative 6; Mismatches 1; Indels 0; Gaps 0;

Db 254 LIRHIPAVIVHG 265





\*Evidence for lateral gene transfer between Archaea and bacteria from  
RT genome sequence of *Thermotoga maritima*.\*

RA Nelson K.E., CLAYTON R.A., GILL S.R., GWINN M.L., DODSON R.J.,  
RA HART D.H., HICKEY E.K., PETERSON J.D., NELSON W.C., KEICHUM K.A.,  
RA McDONALD L., JUTERBACK T.R., MALEK J.A., LINHER K.D., GARRETT M.M.,  
RA STEWART A.M., COTTON M.D., PRATT M.S., PHILLIPS C.A., RICHARDSON D.,  
RA HEIDELBERG S., SUTTON G.G., FLEISCHMANN R.D., WHITE C., SALZBERG S.L.,  
RA SMITH H.O., VENTER J.C., FRASER C.M.,  
RA Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AE001749; AAC35905.1;  
SQ SEQUENCE 190 AA: 22492 MW: B32D72E0 CRC32:

Query Match: 56.4%; Score 53; DB 2; Length 317;  
Best Local Similarity 58.3%; Pred. No. 1.79e-01;  
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

DB 126 FIKRVES:IEEG 137  
|||||  
QY 2 FIKRVSNV:IHG 13

RESULT 15  
ID O66883 PRELIMINARY: PRT: 317 AA.  
AC O66883:  
DT 01-AUG-1998 (TrEMBLrel. 07, Created);  
DI 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)  
DE 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)  
RE THYMIDYLATE SYNTHASE COMPLEMENTING PROTEIN.  
GN THY.  
CS Aquifex aeolicus.  
OC Bacteria: Aquificales: Aquificaceae: Aquifex.  
RN [1]  
RF SEQUENCE FROM N.A.  
SC STRAIN-VFS:  
FX MEDLINE: 98196566.  
RA DECKERT G., WARREN P.V., GAASTERLAND T., YOUNG W.G., LENOX A.L.,  
RA GRAHAM D.E., OVERBECK R., SNEAD M.A., KELLER M., AJJAY M., HUBER R.,  
RA FELDMAN R.A., SHORT J.M., OLSON G.J., SWANSON R.V.;  
RT "The complete genome of the hyperthermophilic bacterium Aquifex  
aeolicus".  
RC Nature 392:353-358(1998).  
RN [2]  
RF SEQUENCE FROM N.A.  
SC STRAIN-VFS:  
RA DECKERT G., WARREN P.V., GAASTERLAND T., YOUNG W.G., LENOX A.L.,  
RA GRAHAM D.E., OVERBECK R., SNEAD M.A., KELLER M., AJJAY M., HUBER R.,  
RA FELDMAN R.A., SHORT J.M., OLSON G.J., SWANSON R.V.;  
PL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AE000700; AAC06847.1;  
SQ SEQUENCE 317 AA: 37542 MW: 6E749CF4 CRC32:

Query Match: 56.4%; Score 53; DB 2; Length 317;  
Best Local Similarity 60.0%; Pred. No. 1.79e-01;  
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

DB 229 VFKKIENLI 238  
|||||  
QY 1 VFIKRVSNVI 10

Search completed: Mon Jun 19 16:16:51 2000  
Job time : 11 secs.

\*\*\*\*\*  
 [WATERMAN] (IX)  
 \*\*\*\*\*

Release 3.1A John F. Collins, BioComputing Research Unit.  
 Copyright (c) 1993-1998 University of Edinburgh, U.K.  
 Distribution rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm  
 Run on: Mon Jun 19 16:05:10 2000; Maspar time 2.77 Seconds  
 Tabular output not generated. 102.539 Million cell updates/sec

Title: >US-09-142-524A-6  
 Description: (1-12) from US09142524A.pap  
 Perfect Score: 72  
 Sequence: 1 ISLKTSGKIAS 12

Scoring table: PAM 150  
 Gap 15

Searched: 18953 seqs, 23686136 residues

Post-processing: Minimum Match 0%  
 Listing first 45 summaries

Database: a-oneseq35  
 i-oneseqp

Statistics: Mean 16.610; Variance 49.150; scale 0.338

Pred. No. is the number of results predicted by chance to have a  
 score greater than or equal to the score of the result being printed.  
 and is derived by analysis of the total score distribution.

## SUMMARIES

| Result<br>No. | Score | Query<br>Match | Length | DB | ID     | Description             | Pred. No. |
|---------------|-------|----------------|--------|----|--------|-------------------------|-----------|
| 1             | 72    | 100.0          | 15     | 1  | R97939 | Japan cedar pollen mat  | 2.46e-01  |
| 2             | 72    | 100.0          | 15     | 1  | W57770 | Residues 341-355 of Cry | 2.46e-01  |
| 3             | 72    | 100.0          | 17     | 1  | R81583 | Cedar pollen allergen   | 2.46e-01  |
| 4             | 72    | 100.0          | 17     | 1  | W80348 | Sugi allergen protein   | 2.46e-01  |
| 5             | 72    | 100.0          | 80     | 1  | W27369 | Multi-epitope peptide   | 2.46e-01  |
| 6             | 72    | 100.0          | 460    | 1  | R69791 | Japanicum allergen res  | 2.46e-01  |
| 7             | 72    | 100.0          | 514    | 1  | R93599 | Japan cedar pollen Cry  | 2.46e-01  |
| 8             | 72    | 100.0          | 514    | 1  | R24333 | Japanese cedar pollen   | 2.46e-01  |
| 9             | 72    | 100.0          | 514    | 1  | R69792 | Japanicum allergen      | 2.46e-01  |
| 10            | 72    | 100.0          | 514    | 1  | R53690 | Japanese cedar pollen   | 2.46e-01  |
| 11            | 72    | 100.0          | 514    | 1  | R81586 | Cedar pollen allergen   | 2.46e-01  |
| 12            | 65    | 90.3           | 11     | 1  | R81577 | Cedar pollen allergen   | 1.67e-00  |
| 13            | 65    | 90.3           | 17     | 1  | R81594 | Cedar pollen allergen   | 1.67e-00  |
| 14            | 61    | 84.7           | 12     | 1  | W14301 | Japanese cedar pollen   | 4.92e-00  |
| 15            | 61    | 84.7           | 15     | 1  | R97940 | Japan cedar pollen mat  | 4.92e-00  |
| 16            | 59    | 81.9           | 20     | 1  | W42197 | 1-cell epitope peptide  | 8.32e-00  |
| 17            | 59    | 81.9           | 514    | 1  | W04346 | Chamaecyparis obtusa p  | 8.32e-00  |
| 18            | 59    | 81.9           | 514    | 1  | W42122 | Japanese cypress polle  | 8.32e-00  |
| 19            | 54    | 75.0           | 35     | 1  | W80342 | Sugi allergen protein   | 3.07e-01  |
| 20            | 54    | 75.0           | 35     | 1  | W80344 | Sugi allergen protein   | 3.07e-01  |
| 21            | 53    | 73.6           | 11     | 1  | W14302 | Japanese cedar pollen   | 3.97e-01  |
| 22            | 53    | 73.6           | 12     | 1  | W14303 | Japanese cedar pollen   | 3.97e-01  |
| 23            | 53    | 73.6           | 33     | 1  | W80341 | Sugi allergen protein   | 3.97e-01  |

|    |    |      |     |   |        |                        |          |
|----|----|------|-----|---|--------|------------------------|----------|
| 24 | 53 | 73.6 | 33  | 1 | W80339 | Sugi allergen protein  | 3.97e-01 |
| 25 | 53 | 73.6 | 33  | 1 | W80340 | Sugi allergen protein  | 3.97e-01 |
| 26 | 53 | 73.6 | 35  | 1 | W80343 | Sugi allergen protein  | 3.97e-01 |
| 27 | 53 | 73.6 | 47  | 1 | W80352 | Sugi allergen protein  | 3.97e-01 |
| 28 | 53 | 73.6 | 47  | 1 | W80357 | Sugi allergen protein  | 3.97e-01 |
| 29 | 53 | 73.6 | 47  | 1 | W80353 | Sugi allergen protein  | 3.97e-01 |
| 30 | 53 | 73.6 | 47  | 1 | W80350 | Sugi allergen protein  | 3.97e-01 |
| 31 | 53 | 73.6 | 47  | 1 | W80351 | Sugi allergen protein  | 3.97e-01 |
| 32 | 53 | 73.6 | 47  | 1 | W80356 | Sugi allergen protein  | 3.97e-01 |
| 33 | 53 | 73.6 | 81  | 1 | W80358 | Sugi allergen protein  | 3.97e-01 |
| 34 | 47 | 65.3 | 572 | 1 | W63312 | Rhizoctonia solani (11 | 1.80e-02 |
| 35 | 47 | 65.3 | 572 | 1 | W60925 | Rhizoctonia solani (11 | 1.80e-02 |
| 36 | 46 | 63.9 | 196 | 1 | P60627 | Beta-urogastrone - bet | 2.30e-02 |
| 37 | 46 | 63.9 | 264 | 1 | W16634 | Beta-lactamase (includ | 2.30e-02 |
| 38 | 46 | 63.9 | 286 | 1 | R22692 | blac31 mutation (2)    | 2.30e-02 |
| 39 | 46 | 63.9 | 286 | 1 | R22702 | blac14 mutation (1)    | 2.30e-02 |
| 40 | 46 | 63.9 | 286 | 1 | R22680 | Modified bla gene prod | 2.30e-02 |
| 41 | 46 | 63.9 | 286 | 1 | R22689 | Modified bla gene prod | 2.30e-02 |
| 42 | 46 | 63.9 | 286 | 1 | R22691 | blac31 mutation (1)    | 2.30e-02 |
| 43 | 46 | 63.9 | 286 | 1 | R22690 | Modified bla gene prod | 2.30e-02 |
| 44 | 46 | 63.9 | 286 | 1 | R22681 | Modified bla gene prod | 2.30e-02 |
| 45 | 46 | 63.9 | 424 | 1 | W97851 | Hypersensitive respons | 2.30e-02 |

## ALIGNMENTS

RESULT 1  
 ID R97939 standard: peptide: 15 AA.  
 AC R97939:  
 DE 16-AUG-1996 (first entry)  
 DE Japan cedar pollen mature allergen Cry j II amino acids 341-355.  
 KW Allergen: epitope; overlapping peptide: Cry j II; cedar pollen;  
 KW Sugi pollinosis; diagnosis: treatment.  
 OS Cryptomeria japonica.  
 PN J08047392.A.  
 PD 20-FEB-1996.  
 PF 07-NOV-1994: 297840.  
 PR 05-NOV-1993: JP-276773.  
 PR 26-MAY-1994: JP-134858.  
 PA (MEIP) MEIJI MILK PROD CO L.  
 DR WPI: 96-166249/17.  
 PT Japan cedar pollen allergen (1, j II) epitope - comprises at least  
 PT part of specified 450 amino acid protein  
 PS Claim 8; Fig 5: 17pp: Japanese.  
 CC R97939:R97960 are overlapping peptides used for the epitope mapping  
 CC of the Japan cedar pollen allergen Cry j II. Cry j II and allergenic  
 CC peptides of it are useful in the diagnosis, prevention and treatment  
 CC of Sugi pollinosis, the allergic reaction to Japan cedar pollen.  
 CC Significant regions of the allergen were identified using the  
 CC overlapping peptides of the full epitope derived from a Cry j II  
 CC antigen-specific T cell line. Amino acids 66-80 (R97884) and 186-200  
 CC (R978908) of the full mature 450 amino acid allergen are the most  
 CC allergenic of the 90 peptides tested.  
 SQ Sequence 15 AA:

Query Match 100.0% Score 72: DB 1: Length 15:  
 Best Local Similarity 100.0%: Pred No. 2.46e-01:  
 Matches 12: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

Db 4 ISLKTSGKIAS 15  
 QY 1 ISLKTSGKIAS 12

RESULT 2  
 ID W57770 standard: peptide: 15 AA.  
 AC W57770:  
 DT 17-SEP-1998 (first entry)  
 DE Residues 341-355 of Cry j 2.  
 KW Cry j 2: Japanese cedar pollen antigen; allergy: immunotherapy.  
 KW HLA class II molecule.  
 OS Cryptomeria japonica.  
 PN WC9820902-A1.

12-MAY-1998.  
 12-NOV-1997: 304129.  
 13-NOV-1996: JP-302053.  
 (MEIP ) MEIJI MILK PROD CO LTD.  
 Dairiki K, Kuno K, Kume A, Sone T;  
 WPI: 98-397617/26.  
 PT Peptides derived from Japanese cedar pollen antigens are  
 immunotherapeutic agents - useful for allergy treatment and typing  
 HLA class II molecules in allergy sufferers  
 Claim 5: Page 53: 50pp: Japanese.  
 This sequence represents residues 341-355 of the Cry j 2 protein, and  
 is a peptide of the invention. The peptides are derived from Japanese  
 cedar pollen antigens, and are used as immunotherapeutic agents in the  
 treatment of allergy. The peptides can be used for identification and  
 typing of the particular HLA class II molecules in an allergy sufferer,  
 and also for peptide immunotherapy of an allergy. Using these peptides,  
 the immunotherapy can be targeted more specifically to the requirements  
 of the individual patient, allowing more effective treatment of an  
 allergy, including those patients for whom treatment with a conventional  
 immunotherapeutic agent is ineffective.  
 Sequence 15 AA:

Query Match 100.0% Score 72: DB 1: Length 15;  
 Best Local Similarity 100.0% Pred. No. 2,46e-01;  
 Matches 12: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

Db 4 ISLKLTSKGKIAS 15  
 |||||  
 QY 1 ISLKLTSKGKIAS 12

## RESULT 3

ID R81583 standard: Peptide: 17 AA.  
 AC R81583:  
 DI 24-MAY-1996 (first entry)  
 DE Cedar pollen allergen peptide 11 (T-cell epitope).  
 KW Cedar; Cryptocarya japonica; pollen; allergen; immunoglobulin E;  
 KW IgE; T-cell epitope; antibody; pollinosis therapy; immunotherapy.  
 OS Synthetic.  
 PN EP-700929-A2.  
 PD 13-MAR-1996.  
 PF 08-SEP-1995: 306295.  
 PR 10-SEP-1994: JP-242137.  
 PR 14-JUN-1993: JP-200221.  
 PR 14-JUN-1993: JP-200204.  
 PA (HAYB ) HAYASHIBARA SEIBUSU KAGAKU.  
 PI Hino K, Saito S, Taniuchi Y;  
 WPI: 96-140976/15.  
 PT New peptide(s) derived from cedar pollen allergens - activate  
 allergen-specific T-cells, but not allergen-specific IgE antibodies.  
 PT used for treating cedar pollinosis  
 Claim 4: Page 29: 36pp: English.  
 CC Synthetic peptides based on portions of cedar pollen allergens A  
 (R81586) and B (R81587) were tested for their ability to activate  
 cedar allergen-specific T-cells, but not allergen-specific IgE  
 antibodies. 6 peptides (R81580-R81585) were identified as T-cell  
 epitopes. These peptides, plus subsequences (R81573-79) essential  
 for T-cell recognition, and homologous peptides (R81588-96) can  
 be used as immunotherapeutic agents to treat or prevent cedar  
 pollinosis, avoiding side-effects such as anaphylaxis.  
 Sequence 17 AA:

Query Match 100.0% Score 72: DB 1: Length 17;  
 Best Local Similarity 100.0% Pred. No. 2,46e-01;  
 Matches 12: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

Db 2 ISLKLTSKGKIAS 13  
 |||||  
 QY 1 ISLKLTSKGKIAS 12

## RESULT 4

ID W80345 standard: peptide: 17 AA.

W80345:  
 DI 11-JAN-1999 (first entry)  
 DE Sugi allergen protein Cryj2 derived epitope for T cells.  
 KW T cell epitope; sugi allergen proteins (Cryj1; Cryj2); treatment;  
 KW sugi-pollinosis; allergic reaction; pollen.  
 OS Synthetic.  
 PN J10259198-A.  
 PD 29-SEP-1998.  
 PF 22-DEC-1997: 353448.  
 PR 24-DEC-1996: JP-343441.  
 PA (HAYB ) HAYASHIBARA SEIBUSU KAGAKU.  
 PA (SANY ) SANKYO CO LTD.  
 DR WPI: 98-577037/49.  
 PT A linked T cell epitope pepi used for the treatment of  
 sugi-pollinosis  
 Claim 7: Page 18: 21pp: Japan.  
 CC W80339-58 represent epitopes T cells derived from the sugi allergen  
 CC proteins Cryj1 (W80339-44, W80339-53 and W80356-58) and Cryj2 (W80345-49  
 CC and W80354-55). The peptides are useful for the treatment of  
 CC sugi-pollinosis, an allergic reaction of the body to pollen.  
 SQ Sequence 17 AA:

Query Match 100.0% Score 72: DB 1: Length 17;  
 Best Local Similarity 100.0% Pred. No. 2,46e-01;  
 Matches 12: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

Db 1 ISLKLTSKGKIAS 12  
 |||||  
 QY 1 ISLKLTSKGKIAS 12

## RESULT 5

ID W27369 standard: peptide: 80 AA.  
 AC W27369:  
 DI 24-MAR-1998 (first entry)  
 DE Multi-epitope peptide used as immunotherapeutic agent #1.  
 KW Multi-epitope peptide; immunotherapeutic agent; allergic disease.  
 KW T-cell epitope region; allergen; lymphocyte; immunoglobulin E.  
 OS Synthetic.  
 PN WO9732600-A1.  
 PD 12-SEP-1997.  
 PF 10-MAR-1997: 300740.  
 PR 10-MAR-1996: JP-080702.  
 PA (MEIP ) MEIJI MILK PROD CO LT.  
 PA Dairiki K, Iwama A, Kuno K, Kume A, Sone T;  
 WPI: 97-479495/43.  
 PT Peptide immunotherapeutic agent to treat allergic diseases  
 contains multi-epitope peptide containing T cell epitope regions  
 from different allergens  
 Claim 6: Page 31: 58pp: Japanese.  
 CC The present sequence represents a multi-epitope peptide which is used as  
 a new immunotherapeutic agent. It comprises T cell epitope regions from 2  
 or more different allergens (preferably linked via arginine or lysine  
 dimers), where the T cell epitope regions have a positive index  
 greater than 100 as measured in a patient group responding to the  
 allergen; have at least 70% reactivity with lymphocytes from patients  
 responding to the allergen; and are not reactive with immunoglobulin E  
 (IgE) antibodies from patients responsive to the allergen. The agent can  
 be used to prevent and treat a wide variety of allergic diseases, e.g. by  
 desensitisation. Side effects, e.g. those mediated by IgE, are reduced.  
 Sequence 80 AA:

Query Match 100.0% Score 72: DB 1: Length 80;  
 Best Local Similarity 100.0% Pred. No. 2,46e-01;  
 Matches 12: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

Db 52 ISLKLTSKGKIAS 53  
 |||||  
 QY 1 ISLKLTSKGKIAS 12

## RESULT 6

ID R6379 standard: Protein: 460 AA.

```

Db      396 ISKLISGKIAS 409
      |||||
QY      1 :SLKLTSGKIAS 12

RESULT      8
ID      R74333 standard; Protein: 514 AA.
AC      R74333:
DE      01-NOV-1995 (first entry)
DI      Japanese cedar pollen allergen.
KW      Japanese cedar; pollen; allergen; allergy; therapy; diagnostic;
KW      desensitize;
KW      Cryptomeria japonica.
PN      EP-655500-A.
PC      31-MAY-1995.
PF      03-NOV-1994; 308117.
PR      05-NOV-1993; JP-299151.
PR      20-DEC-1993; JP-344596.
PR      27-DEC-1993; JP-346814.
PR      (HAYB ) HAYASHIBARA SEIICHIRO KAGAKU.
PI      Kurimoto M, Namba M, Torique K;
PI      WPI: 95-195588/26.
DR      N-PSDB: Q90156.
PT      New Japanese cedar pollen allergen polypeptide - and DNA coding
PT      useful for treatment and diagnosis of cedar pollen allergy

```

CC The gene encoding an allergen of Japanese cedar pollen was isolated  
CC by PCR amplification using primers based on portions of the allergen  
CC protein. The gene was used for recombinant allergen production in  
CC E. coli (vector plasmid pKK-223-3).  
SQ Sequence 514 AA;

CC Seq. (600,000, Premium Pmt. 225)  
SQ Sequence 514 AA;  
Query Match 100.0%; Score 72; EB 1; Length 514;  
Best Local Similarity 100.0%; Pred. No. 2,456-61;

Db 398 ISLKLTSSKIAS 409  
| | | | | | | | | |  
Oy 1 ISLKLTSSKIAS 12

RESULT 5  
ID R65192 standard: Protein: 514 AA.  
AC R65792:  
DE 27-SEP-1995 (first entry)  
KE Japonicum allergen:  
KW IgE cross-linking induced histamine release: anti-allergic peptide:  
KW IgE cross-linking inhibition:  
OS Japonicum sp.  
PN W09592412-A.  
PD 26-JAN-1995.  
PF 15-JUL-1994: J01164.  
PK 16-JUL-1993: JP-177008.  
PR 07-SEP-1993: JP-217725.  
PR 07-SEP-1994: JP-069336.  
PA (NEIP) MELJ1 MILK PROD CO LTD.  
PI Kino K, Kohno Y, Komiyama N, Sone T;  
DR WPI: 95-067159/09.  
DR N-PSDB: Q84045, Q84046.  
PT Peptide anti-allergic agent - Inhibits cross-linking of allergen  
PS with IgE antibody  
PS Example 3: Pages 27-28: 46pp: Japanese.  
CC Q84045 encodes R65792 Japonicum allergen, from which the anti-allergic  
CC peptides R65848-R65909 were derived. The peptides ability to inhibit  
CC the cross-linking of an allergen, to an IgE antibody can be used in  
CC the prevention and treatment of allergic diseases.  
SQ Sequence 514 AA:

```

Query Match      100.0%   Score 72   DB i: Length 514:
Best Local Similarity 100.0%   Pred. NC: 2.40e-01
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0:

Db 398 ISLKLISGKIAS 409

```

Db 398 ISLKLTSCKIAS 409  
1 ISLKLTSCKIAS 12

RESULT 12  
ID R81577 standard: Peptide: 11 AA.

AC R81577 (first entry)  
DE Cedar pollen allergen peptide S.  
KW Cedar; Cryptomeria japonica; pollen; allergen; immunoglobulin E  
OS IgE antibody; pollinosis; therapy; immunotherapy.  
PN EP-70029-A2.  
PR 13-MAR-1996: 306295.  
PR 08-SEP-1995: JP-242137.  
PR 10-SEP-1994: JP-242137.  
PR 14-JUL-1995: JP-200221.  
PR 14-JUL-1995: JP-200204.  
PA (HAYB.) HAYASHIBARA SEIBUTSU KAGAKU.  
PI Hino K, Saito S, Taniuchi Y.  
PI WPI: 96-140976/15.  
DR New peptide(s) derived from cedar pollen allergens - activate  
PS Claim 2: Fig 4: 9pp: English.  
CC The sequence is of a Japanese cedar pollen allergen Cry j  
CC 11: the protein and its fragments can be used for diagnosis and  
CC treatment of Japanese cedar pollinosis and to identify similar  
CC sequences in other plants.  
CC See also R81586-6.  
CC Sequence 514 AA.

Query Match 100.0% Score 72: DB 1: Length 514;

Best Local Similarity 100.0% Pred. No. 2.46e-01; Mismatches 0; Indels 0; Gaps 0;

Db 398 ISLKLTSCKIAS 409  
1 ISLKLTSCKIAS 12

RESULT 11  
ID R81586 standard: Protein: 514 AA.

AC R81586 (first entry)  
DE Cedar pollen allergen A.  
KW Cedar; pollen; allergen; immunoglobulin E; IgE; T-cell epitope;  
OS Cryptomeria japonica; antibody; pollinosis; therapy; immunotherapy.  
PN EP-70029-A2.  
PR 13-MAR-1996: 306295.  
PR 08-SEP-1995: 306295.  
PR 10-SEP-1994: JP-242137.  
PR 14-JUL-1995: JP-200221.  
PR 14-JUL-1995: JP-200204.  
PA (HAYB.) HAYASHIBARA SEIBUTSU KAGAKU.  
PI Hino K, Saito S, Taniuchi Y.  
PI WPI: 96-140976/15.  
DR New peptide(s) derived from cedar pollen allergens - activate  
PS Claim 5: Page 29-30: 36pp: English.  
CC Synthetic peptides based on portions of cedar pollen allergens A  
CC (R81586) and B (R81587) were tested for their ability to activate  
CC cedar allergen-specific T-cells, but not allergen-specific IgE  
CC antibodies. 6 peptides (R81580-R81585) were identified as T-cell  
CC epitopes. These peptides, plus subsequences (R81573-79) essential  
CC for T-cell recognition, and homologous peptides (R81588-95) can  
CC be used as immunotherapeutic agents to treat or prevent cedar  
CC pollinosis, avoiding side-effects such as anaphylaxis.  
CC Sequence 514 AA.

Query Match 100.0% Score 72: DB 1: Length 514;

Best Local Similarity 100.0% Pred. No. 2.46e-01; Mismatches 0; Indels 0; Gaps 0;

Db 398 ISLKLTSCKIAS 409  
1 ISLKLTSCKIAS 12

RESULT 12

ID R81577 standard: Peptide: 11 AA.  
AC R81577 (first entry)  
DE Cedar pollen allergen peptide S.  
KW Cedar; Cryptomeria japonica; pollen; allergen; immunoglobulin E  
OS IgE antibody; pollinosis; therapy; immunotherapy.  
PN EP-70029-A2.  
PR 13-MAR-1996: 306295.  
PR 08-SEP-1995: JP-242137.  
PR 10-SEP-1994: JP-242137.  
PR 14-JUL-1995: JP-200221.  
PR 14-JUL-1995: JP-200204.  
PA (HAYB.) HAYASHIBARA SEIBUTSU KAGAKU.  
PI Hino K, Saito S, Taniuchi Y.  
PI WPI: 96-140976/15.  
DR New peptide(s) derived from cedar pollen allergens - activate  
PS Claim 3: Page 27: 36pp: English.  
CC Synthetic peptides based on portions of cedar pollen allergens A  
CC (R81586) and B (R81587) were tested for their ability to activate  
CC cedar allergen-specific T-cells, but not allergen-specific IgE  
CC antibodies. 6 peptides (R81580-R81585) were identified as T-cell  
CC epitopes. These peptides, plus subsequences (R81573-79) essential  
CC for T-cell recognition, and homologous peptides (R81588-95) can  
CC be used as immunotherapeutic agents to treat or prevent cedar  
CC pollinosis, avoiding side-effects such as anaphylaxis.  
CC Sequence 11 AA.

Query Match 100.0% Score 65: DB 1: Length 11;

Best Local Similarity 100.0% Pred. No. 1.67e-00; Mismatches 0; Indels 0; Gaps 0;

Db 1 ISLKLTSCKIAS 11  
2 ISLKLTSCKIAS 12

RESULT 11

ID R81544 standard: Peptide: 17 AA.  
AC R81544 (first entry)  
DE Cedar pollen allergen peptide 22 (therologue)  
KW Cedar; Cryptomeria japonica; pollen; allergen; immunoglobulin E  
OS IgE; T-cell epitope; antibody; pollinosis; therapy; immunotherapy.  
PN EP-70029-A2.  
PR 13-MAR-1996: 306295.  
PR 08-SEP-1995: 306295.  
PR 10-SEP-1994: JP-242137.  
PR 14-JUL-1995: JP-200221.  
PR 14-JUL-1995: JP-200204.  
PA (HAYB.) HAYASHIBARA SEIBUTSU KAGAKU.  
PI Hino K, Saito S, Taniuchi Y.  
PI WPI: 96-140976/15.  
DR New peptide(s) derived from cedar pollen allergens - activate  
PS Claim 8: Page 33: 36pp: English.  
CC Synthetic peptides based on portions of cedar pollen allergens A  
CC (R81586) and B (R81587) were tested for their ability to activate  
CC cedar allergen-specific T-cells, but not allergen-specific IgE  
CC antibodies. 6 peptides (R81580-R81585) were identified as T-cell  
CC epitopes. These peptides, plus subsequences (R81573-79) essential  
CC for T-cell recognition, and homologous peptides (R81588-95)

CC that contain 1 or more amino acid substitution(s) can be used  
 CC as immunotherapeutic agents to treat or prevent cedar pollinosis,  
 CC avoiding side-effects such as anaphylaxis.

SQ Sequence 17 AA:

Query Match 90.3%; Score 65; DB 1; Length 17;

Best Local Similarity 100.0%; Pred. No. 1.97e+00;

Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 3 SLKLTSGKIAS 13

QY 2 SLKLTSGKIAS 12

#### RESULT 14

ID W14301 standard; peptide: 12 AA.

AC W14301

DE 30-APR-1997 (first entry)

DE Japanese cedar pollen allergen-derived peptide 20.

KW Japanese cedar; sugi; pollen; pollinosis; pollenosis; allergen;

KW allergy; T-cell; epitope; TCR; T-cell receptor; activation;

KW immune tolerance; Cry1.

CS Cryptomeria japonica.

PS Synthetic.

PN J08334391-A.

PE 17-DEC-1996.

PF 18-JUL-1995; 181438.

PP 07-APR-1995; JP-082519.

FA (HAYB ) HAYASHIBARA SEIBUTSU KAGAKU.

FA (SANY ) SANKYO CO LTD.

DE WPI: 97-095487/09.

P1 Peptide allergen derived from Japanese cedar pollen - causes T cell

P1 response specific to cedar pollen, for treatment of pollinosis

PS Claim 29; Page 20; 21pp; Japanese.

CC The present sequence is one of 24 claimed peptides which were

CC synthesised based on Japanese cedar pollen sequences. This peptide

CC was shown to have Japanese cedar pollen antigen T cell epitope

CC activity by using T cells isolated from a cedar pollinosis patient.

CC The peptide produces little or no anaphylaxis.

SQ Sequence 12 AA:

Query Match 94.7%; Score 51; DB 1; Length 12;

Best Local Similarity 100.0%; Pred. No. 4.90e+00;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 1 LKLTSGKIAS 10

QY 3 LKLTSGKIAS 12

#### RESULT 15

ID R97940 standard; peptide: 15 AA.

AC R97940:

DE 16-AUG-1996 (first entry)

DE Japan cedar pollen mature allergen Cry j II amino acids 346-360.

KW Allergen; epitope; overlapping peptide; Cry j II; cedar pollen;

KW Sugi pollinosis; diagnosis; treatment.

CS Cryptomeria japonica.

PN J08047392-A.

PE 20-FEB-1996.

PF 07-NOV-1994; 297840.

PP 05-NOV-1993; JP-276773.

PR 25-MAY-1994; JP-134858.

PA (MEIP ) MELIJI MILK PROD CO LTD.

DE WPI: 96-166249/17.

P1 Japan cedar pollen allergen Cry j II epitope - comprises at least

P1 part of specified 460 amino acid protein

PS Claim 8; Fig 5; 17pp; Japanese.

CC R97871-R97960 are overlapping peptides used for the epitope mapping

CC of the Japan cedar pollen allergen Cry j II. Cry j II and allergenic

CC peptides of it are useful in the diagnosis, prevention and treatment

CC of Sugi pollinosis, the allergic reaction to Japan cedar pollen.

CC Significant regions of the allergen were identified using the

CC overlapping peptides of the full epitope derived from a Cry j II  
 CC antigen-specific T cell line. Amino acids 65-80 (R97884) and 181-200  
 CC (R978908) of the full mature 460 amino acid allergen are the most  
 CC allergenic of the 90 peptides tested.

SQ Sequence 15 AA:

Query Match 84.7%; Score 51; DB 1; Length 15;

Best Local Similarity 100.0%; Pred. No. 4.90e+00;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 1 LKLTSGKIAS 10

QY 3 LKLTSGKIAS 12

Search completed: Mon Jun 19 16:05:21 2000

Job time : 11 secs.

\*\*\*\*\*  
 [W][A][T][E][R][M][A][N]  
 [S][E][Q][U][E][N][C][E]  
 [I][N][F][O][R][M][A][T][I][O][N]  
 [I][M]  
 \*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
 Copyright (c) 1993-1998 University of Edinburgh, U.K.  
 Distribution rights by Oxford Molecular Ltd

Mfsrcch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Tue Jun 20 13:34:47 2000; Maspar time 2.23 seconds

Tabular output not generated. 77.802 Million cell updates/sec

Title: >US-09-142-524A-6  
 Description: (1-12) from US09142524A.pep  
 Perfect Score: 72  
 Sequence: 1 ISLKLTSQKIAS 12

Scoring table: PWM 150  
 Gap 15

Searched: 145341 seqs, 14437480 residues

Post-processing: Minimum Match 0%  
 Listing first 45 summaries

Database: a-issued  
 1:5A\_COMB 2:5B\_COMB 3:5\_COI 4:PCT\_COMB 5:backfiles:

Statistics: Mean 15.504; Variance 46.068; scale 0.337

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARY:

| Result No. | Score | Query Match | Length | ID | Description                      | Pred. No. |
|------------|-------|-------------|--------|----|----------------------------------|-----------|
| 1          | 47    | 65.3        | 572    | 2  | US-08-993-Sequence 7, Applicatio | 8.72e+01  |
| 2          | 47    | 65.3        | 572    | 2  | US-09-032-Sequence 7, Applicatio | 8.72e+01  |
| 3          | 46    | 63.9        | 263    | 1  | US-08-407-Sequence 2, Applicatio | 1.12e+02  |
| 4          | 46    | 63.9        | 284    | 2  | US-08-719-Sequence 8, Applicatio | 1.12e+02  |
| 5          | 46    | 63.9        | 285    | 2  | US-08-719-Sequence 2, Applicatio | 1.12e+02  |
| 6          | 46    | 63.9        | 285    | 2  | US-08-719-Sequence 6, Applicatio | 1.12e+02  |
| 7          | 46    | 63.9        | 286    | 1  | US-08-346-Sequence 16, Applicati | 1.12e+02  |
| 8          | 46    | 63.9        | 286    | 2  | US-08-719-Sequence 4, Applicatio | 1.12e+02  |
| 9          | 46    | 63.9        | 286    | 1  | US-07-721-Sequence 2, Applicatio | 1.12e+02  |
| 10         | 46    | 63.9        | 286    | 4  | PCT-US91-0Sequence 16, Applicati | 1.12e+02  |
| 11         | 46    | 63.9        | 286    | 1  | US-08-339-Sequence 2, Applicatio | 1.12e+02  |
| 12         | 46    | 63.9        | 359    | 1  | US-08-503-Sequence 2, Applicatio | 1.12e+02  |
| 13         | 46    | 63.9        | 359    | 2  | US-08-576-Sequence 2, Applicatio | 1.12e+02  |
| 14         | 46    | 63.9        | 359    | 3  | US-08-899-Sequence 2, Applicatio | 1.12e+02  |
| 15         | 46    | 63.9        | 359    | 2  | US-08-972-Sequence 2, Applicatio | 1.12e+02  |
| 16         | 46    | 63.9        | 1186   | 4  | PCT-US93-1Sequence 6, Applicatio | 1.12e+02  |
| 17         | 46    | 63.9        | 1186   | 1  | US-08-357-Sequence 6, Applicatio | 1.12e+02  |
| 18         | 46    | 63.9        | 1186   | 2  | US-08-590-Sequence 4, Applicatio | 1.12e+02  |
| 19         | 46    | 63.9        | 1186   | 1  | US-08-485-Sequence 4, Applicatio | 1.12e+02  |
| 20         | 46    | 63.9        | 1186   | 2  | US-09-184-Sequence 4, Applicatio | 1.12e+02  |
| 21         | 46    | 63.9        | 1308   | 2  | US-08-996-Sequence 2, Applicatio | 1.12e+02  |
| 22         | 45    | 62.5        | 331    | 4  | PCT-US95-1Sequence 4, Applicatio | 1.44e+02  |
| 23         | 45    | 62.5        | 331    | 1  | US-08-208-Sequence 8, Applicatio | 1.44e+02  |

# ALIGNMENTS

RESULT 1  
 ID US-08-993-318A-7 STANDARD: PRT: 572 AA.  
 XX  
 AC xxxxxx  
 XX  
 DT  
 DE Sequence 7, Application US/08993318A  
 XX  
 CC Sequence 7, Application US/08993318A  
 CC Patent No. 5998353  
 CC GENERAL INFORMATION:  
 CC APPLICANT: Pedersen, Anders  
 CC APPLICANT: Svendsen, Allan  
 CC APPLICANT: Schneider, Palie  
 CC APPLICANT: Rasmussen, G. the  
 CC APPLICANT: Cherry, Joel  
 CC TITLE OF INVENTION: LAC USE MUTANTS  
 CC NUMBER OF SEQUENCES: 10  
 CC CORRESPONDENCE ADDRESS:  
 CC ADDRESSEE: NO. 5998353 NO. 5998353disk of NO. 5998353, AMERICA  
 CC STREET: 405 Lexington Avenue  
 CC CITY: New York  
 CC COUNTRY: USA  
 CC ZIP: 10174  
 CC COMPUTER READABLE FORM:  
 CC MEDIUM TYPE: Floppy disk  
 CC COMPUTER: IBM PC compatible  
 CC OPERATING SYSTEM: PC-DOS/MS-DOS  
 CC SOFTWARE: Patent Release #1.0, Version #1.30 (EPC)  
 CC CURRENT APPLICATION DATA:  
 CC APPLICATION NUMBER: US/08/993,318A  
 CC FILING DATE: December 18, 1997  
 CC CLASSIFICATION: 435  
 CC ATTORNEY/AGENT INFORMATION:  
 CC NAME: Gregq, Valeta A.  
 CC REGISTRATION NUMBER: 33,728  
 CC REFERENCE/DOCKET NUMBER: 5032.200-US  
 CC TELECOMMUNICATION INFORMATION:  
 CC TELEPHONE: 212-867-012  
 CC TELEFAX: 212-878-9655  
 CC TELEX:  
 CC INFORMATION FOR SEQ ID NO: 7:  
 CC SEQUENCE CHARACTERISTICS:  
 CC LENGTH: 572 amino acids  
 CC TYPE: amino acid  
 CC STRANDEDNESS: single



[illegible]



Matches 7: Conservative 0: Mismatches 3: Indels 0: Gaps 0:

Db 24 IELDLSGK: 33  
 QY 1 ISLKLTSK: 10

RESULT 7  
 ID US-08-346-333-16 STANDARD: PRT: 286 AA.  
 XX XXXXX  
 DE

Sequence 16, Application US/08346333  
 Patent No. 597153  
 GENERAL INFORMATION:  
 APPLICANT: Botstein, David  
 TITLE OF INVENTION: Methods for modifying DNA and for  
 TITLE OF INVENTION: detecting effects of such modification on interaction of  
 TITLE OF INVENTION: encoded modified polypeptides with target substrates.  
 NUMBER OF SEQUENCES: 86  
 CORRESPONDENCE ADDRESS:  
 ADDRESS: Richard F. Treccartin  
 STREET: 4 Embarcadero Center, Suite 3400  
 CITY: San Francisco  
 STATE: California  
 COUNTRY: USA  
 ZIP: 94111

COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/346,333  
 FILING DATE:  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US/08/039,501  
 FILING DATE:  
 APPLICATION NUMBER: US/07/602,156  
 FILING DATE: 22-OCT-1999  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Treccartin, Richard F.  
 REGISTRATION NUMBER: 31,801  
 REFERENCE/DOCKET NUMBER: A 53469/RP  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (415) 781-1989  
 TELEFAX: (415) 398-3245  
 INFORMATION FOR SEQ ID NO: 16:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 286 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 SEQUENCE 286 AA: 31456 MW: 397468 CN:

Query Match: 63.9% Score 46: DB 1: Length 286:  
 Best Local Similarity 70.0%: Pct. Ident. 0:  
 Matches 7: Conservative 0: Mismatches 3: Indels 0: Gaps 0:

Db 45 IELDLSGK: 54  
 QY 1 ISLKLTSK: 10

RESULT 8  
 ID US-08-719-697-4 STANDARD: PRT: 286 AA.  
 XX XXXXX  
 DE

Sequence 16, Application US/08346333  
 Patent No. 597153  
 GENERAL INFORMATION:  
 APPLICANT: Botstein, David  
 TITLE OF INVENTION: Methods for modifying DNA and for  
 TITLE OF INVENTION: detecting effects of such modification on interaction of  
 TITLE OF INVENTION: encoded modified polypeptides with target substrates.  
 NUMBER OF SEQUENCES: 86  
 CORRESPONDENCE ADDRESS:  
 ADDRESS: Richard F. Treccartin  
 STREET: 4 Embarcadero Center, Suite 3400  
 CITY: San Francisco  
 STATE: California  
 COUNTRY: USA  
 ZIP: 94111

COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/346,333  
 FILING DATE:  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US/08/039,501  
 FILING DATE:  
 APPLICATION NUMBER: US/07/602,156  
 FILING DATE: 22-OCT-1999  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Treccartin, Richard F.  
 REGISTRATION NUMBER: 31,801  
 REFERENCE/DOCKET NUMBER: A 53469/RP  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (415) 781-1989  
 TELEFAX: (415) 398-3245  
 INFORMATION FOR SEQ ID NO: 16:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 286 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 SEQUENCE 286 AA: 31456 MW: 397468 CN:

xxxxxx

Sequence 4, Application US/08719697  
 Patent No. 5928888  
 GENERAL INFORMATION:  
 APPLICANT: Whitney, Michael A.  
 TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR SENSITIVE  
 TITLE OF INVENTION: AND RAPID, FUNCTIONAL IDENTIFICATION OF GENOMIC  
 TITLE OF INVENTION: POLYNUCLEOTIDES AND SECONDARY SCREENING CAPABILITIES  
 NUMBER OF SEQUENCES: 10  
 CORRESPONDENCE ADDRESS:  
 ADDRESS: Fish & Richardson P.C.  
 STREET: 4225 Executive Square, Suite 1400  
 CITY: La Jolla  
 STATE: CA  
 COUNTRY: USA  
 ZIP: 92037

COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/719,697  
 FILING DATE: 26-SEP-1996  
 CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Haile, Lisa A.  
 REGISTRATION NUMBER: 38,347  
 REFERENCE/DOCKET NUMBER: 04366/00001  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 619/678-5076  
 TELEFAX: 619/678-5099  
 INFORMATION FOR SEQ ID NO: 4:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 286 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 SEQUENCE 286 AA: 31570 MW: 346420 CN:

Query Match: 63.9% Score 46: DB 1: Length 286:  
 Best Local Similarity 70.0%: Pct. Ident. 0:  
 Matches 7: Conservative 0: Mismatches 3: Indels 0: Gaps 0:

Db 45 IELDLSGK: 54  
 QY 1 ISLKLTSK: 10

RESULT 9  
 ID US-07-721-775A-2 STANDARD: PRT: 286 AA.  
 XX XXXXX  
 DE

Sequence 2, Application US/07721775A  
 Patent No. 5180666  
 GENERAL INFORMATION:  
 APPLICANT: States, J. Christopher  
 APPLICANT: Hines, Ronald N.  
 TITLE OF INVENTION: NO. 5180666ak, Raymond F.  
 TITLE OF INVENTION: METHOD AND CELL LINE FOR TESTING  
 TITLE OF INVENTION: MUTAGENICITY OF A CHEMICAL  
 NUMBER OF SEQUENCES: 2  
 CORRESPONDENCE ADDRESS:





CC Patent No. 6020201  
CC GENERAL INFORMATION:  
CC APPLICANT: Gerardy-Schahn, Rita; Fukuda, Minoru;  
CC APPLICANT: Nakayama, Jun; Eckhardt, Matthias  
CC TITLE OF INVENTION: Isolated Polysialyl Transferases.  
CC TITLE OF INVENTION: Nucleic Acid Molecules Coding Therefor, Methods of  
CC TITLE OF INVENTION: Production and Use  
CC NUMBER OF SEQUENCES: 8  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Felfe & Lynch  
CC STREET: 805 Third Avenue  
CC CITY: New York City  
CC STATE: New York  
CC COUNTRY: USA  
CC ZIP: 10022  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Diskette, 3.5 inch, 360 kb storage  
CC COMPUTER: IBM PS/2  
CC OPERATING SYSTEM: PC-DOS  
CC SOFTWARE: Wordperfect  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/899,545  
CC FILING DATE:  
CC CLASSIFICATION:  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: 08/503,133  
CC FILING DATE:  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Hanson, No. 5020201man D.  
CC REGISTRATION NUMBER: 30,945  
CC REFERENCE/DOCKET NUMBER: BOER 1050  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (212) 688-9200  
CC TELEFAX: (212) 838-3884  
CC INFORMATION FOR SEQ ID NO: 2:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 359 amino acids  
CC TYPE: amino acid  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: protein  
CC SEQUENCE 359 AA; 41227 MW; 690704 CN;  
  
Query Match 63.9%; Score 46; DB 3; Length 359;  
Best Local Similarity 75.0%; Pred. NO. 1.12e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Ep 348 AKLTGK 355  
QY 2 SKLTSGK 9  
  
RESULT 15  
ID US-08-972-498-2 STANDARD: PRT: 359 AA.  
XX  
XX AC xxxxxx  
XX  
XX DT  
XX  
XX DE Sequence 2, Application US/08972498  
XX Sequence 2, Application US/08972498  
XX Patent No. 5959078  
XX GENERAL INFORMATION:  
XX APPLICANT: Gerardy-Schahn, Rita; Fukuda, Minoru;  
XX APPLICANT: Nakayama, Jun; Eckhardt, Matthias  
XX TITLE OF INVENTION: Isolated Polysialyl Transferases,  
XX TITLE OF INVENTION: Nucleic Acid Molecules Coding Therefor, Methods of  
XX TITLE OF INVENTION: Production and Use  
XX NUMBER OF SEQUENCES: 8  
XX CORRESPONDENCE ADDRESS:  
XX ADDRESSEE: Felfe & Lynch  
XX STREET: 805 Third Avenue  
XX CITY: New York City

CC STATE: New York  
CC COUNTRY: USA  
CC ZIP: 10022  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Diskette, 3.5 inch, 360 kb storage  
CC COMPUTER: IBM PS/2  
CC OPERATING SYSTEM: PC-DOS  
CC SOFTWARE: Wordperfect  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/972,498  
CC FILING DATE:  
CC CLASSIFICATION: 536  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: 08/576,775  
CC FILING DATE: 21-December-1995  
CC APPLICATION NUMBER: 08/503,133  
CC FILING DATE: 17-July-1995  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: PC/EP94/04289  
CC FILING DATE: 22-December-1994  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Hanson, No. 5959078man D.  
CC REGISTRATION NUMBER: 30,945  
CC REFERENCE/DOCKET NUMBER: BOER 1050.1  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (212) 688-9200  
CC TELEFAX: (212) 838-3884  
CC INFORMATION FOR SEQ ID NO: 1:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 359 amino acid  
CC TYPE: amino acid  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: protein  
CC SEQUENCE 359 AA; 41227 MW; 690704 CN;  
  
Query Match 63.9%; Score 46; DB 2; Length 359;  
Best Local Similarity 75.0%; Pred. NO. 1.12e+02;  
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
  
Db 348 AKLTGK 355  
QY 2 SKLTSGK 9  
  
Search completed: Tue Jun 20 13:34:52 2000  
Job time: 5 secs.

\*\*\*\*\*  
 W O R L D  
 \*\*\*\*\*  
 (TW)

Release 3.1A John F. Collins, BioComputing Research Unit.  
 Copyright (c) 1993-1998 University of Edinburgh, U.K.  
 Distribution rights by Oxford Molecular Ltd

MFASTCH\_PP protein - protein database search, using Smith-Waterman algorithm

Run on: Mon Jun 19 16:23:50 2000; MasPar time 14.77 seconds  
 Tabular output not generated. 81.852 Million cell updates/sec

Title: >US-09-142-524A-6  
 Description: (1-12) from US09142524A.pep  
 Perfect Score: 72  
 Sequence: 1 ISLK:TSCKIAS 12

Scoring table: PAM 150  
 Gap 15

Searched: 721208 seqs, 100765575 residues

Post-processing: Minimum Match 0%  
 Listing first 45 summaries

Database: a-repding

1:PCI 2:06 3:060 4:07 5:080 6:051 7:082 8:083 9:084A  
 10:084B 11:085 12:086 13:087 14:088 15:089 16:090 17:091  
 18:092 19:093 20:094 21:095 22:NEWP 23:NEW5C 24:NEW08  
 25:NEW09

Statistics: Mean 18.571; Variance 42.966; scale 0.432

Pred. No. is the number of results predicted by chance to have a  
 score greater than or equal to the score of the result being printed,  
 and is derived by analysis of the total score distribution.

# SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description                       | Pred. No. |
|------------|-------|-------------|--------|-------|-----------------------------------|-----------|
| 1          | 72    | 100.0       | 17     | 11    | US-08-526- Sequence 11, Applicati | 1.63e-01  |
| 2          | 72    | 100.0       | 24     | 10    | US-08-467- Sequence 192, Applicat | 1.63e-01  |
| 3          | 72    | 100.0       | 24     | 10    | US-08-467- Sequence 192, Applicat | 1.63e-01  |
| 4          | 72    | 100.0       | 24     | 10    | US-08-468- Sequence 192, Applicat | 1.63e-01  |
| 5          | 72    | 100.0       | 24     | 7     | US-08-226- Sequence 192, Applicat | 1.63e-01  |
| 6          | 72    | 100.0       | 24     | 8     | US-08-350- Sequence 192, Applicat | 1.63e-01  |
| 7          | 72    | 100.0       | 24     | 10    | US-08-467- Sequence 192, Applicat | 1.63e-01  |
| 8          | 72    | 100.0       | 80     | 17    | US-08-142- Sequence 190, Applicat | 1.63e-01  |
| 9          | 72    | 100.0       | 127    | 10    | US-08-350- Sequence 190, Applicat | 1.63e-01  |
| 10         | 72    | 100.0       | 127    | 8     | US-08-467- Sequence 190, Applicat | 1.63e-01  |
| 11         | 72    | 100.0       | 127    | 10    | US-08-467- Sequence 190, Applicat | 1.63e-01  |
| 12         | 72    | 100.0       | 127    | 7     | US-08-226- Sequence 190, Applicat | 1.63e-01  |
| 13         | 72    | 100.0       | 127    | 10    | US-08-467- Sequence 190, Applicat | 1.63e-01  |
| 14         | 72    | 100.0       | 127    | 10    | US-08-468- Sequence 190, Applicat | 1.63e-01  |
| 15         | 72    | 100.0       | 460    | 11    | US-08-571- Sequence 1, Applicatio | 1.63e-01  |
| 16         | 72    | 100.0       | 460    | 15    | US-08-980- Sequence 1, Applicatio | 1.63e-01  |
| 17         | 72    | 100.0       | 514    | 15    | US-08-980- Sequence 2, Applicatio | 1.63e-01  |
| 18         | 72    | 100.0       | 514    | 4     | US-07-975- Sequence 2, Applicatio | 1.63e-01  |
| 19         | 72    | 100.0       | 514    | 11    | US-08-526- Sequence 14, Applicati | 1.63e-01  |
| 20         | 72    | 100.0       | 514    | 8     | US-08-350- Sequence 134, Applicat | 1.63e-01  |

21 72 100.0 514 10 US-08-467- Sequence 134, Applicat 1.63e-01 |

22 72 100.0 514 10 US-08-468- Sequence 134, Applicat 1.63e-01 |

23 72 100.0 514 11 US-08-571- Sequence 2, Applicatio 1.63e-01 |

24 72 100.0 514 10 US-08-467- Sequence 134, Applicat 1.63e-01 |

25 72 100.0 514 10 US-08-467- Sequence 134, Applicat 1.63e-01 |

26 72 100.0 514 8 US-08-334- Sequence 2, Applicatio 1.63e-01 |

27 72 100.0 514 7 US-08-226- Sequence 134, Applicat 1.63e-01 |

28 72 100.0 514 1 PCT-US93-1 Sequence 2, Applicatio 1.63e-01 |

29 65 90.3 11 11 US-08-526- Sequence 5, Applicatio 1.63e-00 |

30 65 90.3 17 11 US-08-526- Sequence 22, Applicati 1.63e-00 |

31 54 75.0 35 3 US-03-172- Sequence 5, Applicatio 1.63e-01 |

32 52 72.2 74 3 US-60-169- Sequence 4624, Applicat 1.64e-02 |

33 52 72.2 74 3 US-60-160- Sequence 3284, Applicat 1.64e-02 |

34 52 72.2 100 3 US-60-162- Sequence 2714, Applicat 1.64e-02 |

35 52 72.2 100 3 US-60-169- Sequence 5348, Applicat 1.64e-02 |

36 50 59.4 199 20 US-09-417- Sequence 25733, Applic 1.9e-02 |

37 49 68.1 298 20 US-09-417- Sequence 38423, Applic 2.55e-02 |

38 49 68.1 394 17 US-09-107- Sequence 5184, Applicat 2.55e-02 |

39 49 68.1 573 20 US-09-450- Sequence 5090, Applicat 2.55e-02 |

40 49 68.1 1001 3 US-60-173- Sequence 6781, Applicat 2.55e-02 |

41 49 68.1 1001 3 US-60-167- Sequence 8461, Applicat 2.55e-02 |

42 48 66.7 77 3 US-01-162- Sequence 2847, Applicat 3.45e-02 |

43 48 66.7 253 25 US-07-452- Sequence 8265, Applicat 3.45e-02 |

44 48 66.7 318 20 US-07-417- Sequence 25464, Applicat 3.45e-02 |

45 48 66.7 344 3 US-60-128- Sequence 3273, Applicat 3.45e-02 |

## ALIGNMENTS

RESULT 1  
 ID US-08-526-179-11 STANDARD: PRT: 17 AA.

XX  
 AC xxxxxx

DE Sequence 11, Application US/08526179

XX Sequence 11, Application US/08 26179

CC GENERAL INFORMATION:

CC APPLICANT: SAITO, Saburo

CC APPLICANT: HINO, Katsuhiko

CC APPLICANT: TANIGUCHI, Yoshitomi

CC APPLICANT: KURIMOTO, Masahiko

CC TITLE OF INVENTION: METHOD AND USES OF THE SAME

CC NUMBER OF SEQUENCES: 24

CC CORRESPONDENCE ADDRESS:

CC ADDRESSES: BROWDY AND NEIMARK

CC STREET: 419 Seventh Street, N.W., Suite 300

CC CITY: Washington

CC STATE: D.C.

CC COUNTRY: USA

CC ZIP: 20004

CC COMPUTER READABLE FORM:

CC MEDIUM TYPE: Floppy disk

CC COMPUTER: IBM PC Compatible

CC OPERATING SYSTEM: PC-DOS/MS-DOS

CC SOFTWARE: Patentin Release #1.0, Version #1.30

CC CURRENT APPLICATION DATA:

CC APPLICATION NUMBER: US/08/526.179

CC FILING DATE:

CC CLASS/FICATION: 530

CC PRIOR APPLICATION DATA:

CC APPLICATION NUMBER: JP 242,137/1994

CC FILING DATE: 10-SEP-1994

CC PRIOR APPLICATION DATA:

CC APPLICATION NUMBER: JP 200,204/1995

CC FILING DATE: 14-JUL-1995

CC PRIOR APPLICATION DATA:

CC APPLICATION NUMBER: JP 200,221/1995

CC FILING DATE: 14-JUL-1995

CC ATTORNEY/AGENT INFORMATION:

CC NAME: BROWDY, Roger L.

CC REGISTRATION NUMBER: 25,619  
 CC REFERENCE/DOCKET NUMBER: SAITO-19  
 CC TELECOMMUNICATION INFORMATION:  
 CC TELEPHONE: 202-628-5197  
 CC TELEFAX: 202-737-3528  
 CC TELEX: 2486633

## INFORMATION FOR SEQ ID NO: 11:

## SEQUENCE CHARACTERISTICS:

LENGTH: 17 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide

SEQUENCE: 17 AA: 1778 MW: 1362 CN:

Query Match 100.0% Score 72: DB 11: Length 17:

Best Local Similarity 100.0%: Pred. No. 1,63e-01:

Matches 12: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

CC 2 ISLKLTSOKIAS 13

CC 1 ISLKLTSOKIAS 12

## RESULT 2

CC ID US-08-467-697-192 STANDARD: PRT: 24 AA.

CC AC XXXXX

CC DJ

CC XX

CC XX

CC XX

CC XX

CC XX

CC XX

CC XX

CC XX

CC XX

CC XX

CC XX

CC XX

CC XX

CC XX

CC XX

CC XX

CC XX

CC XX

CC XX

CC XX

CC XX

CC XX

CC XX

CC XX

CC XX

CC XX

CC XX

CC XX

CC XX

CC XX

CC XX

CC XX

CC XX

CC XX

CC XX

CC XX

CC XX

CC XX

CC XX

CC XX

CC XX

CC XX

CC XX

CC XX

CC XX

CC XX

## INFORMATION FOR SEQ ID NO: 192:

## SEQUENCE CHARACTERISTICS:

LENGTH: 24 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: peptide

FRAGMENT TYPE: Internal

SEQUENCE: 24 AA: 2551 MW: 3329 CN:

Query Match 100.0% Score 72: DB 10: Length 24:

Best Local Similarity 100.0%: Pred. No. 1,61e-01:

Matches 12: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

CC 13 ISLKLTSOKIAS 24

CC 1 ISLKLTSOKIAS 12

## RESULT 3

CC ID US-08-467-006-192 STANDARD: PRT: 24 AA.

CC AC XXXXX

CC DJ

CC XX

CC XX

CC XX

CC XX

CC XX

CC XX

CC XX

CC XX

CC XX

CC XX

CC XX

CC XX

CC XX

CC XX

CC XX

CC XX

CC XX

CC XX

CC XX

CC XX

CC XX

CC XX

CC XX

CC XX

CC XX

CC XX

CC XX

CC XX

CC XX

CC XX

CC XX

CC XX

CC XX

CC XX

CC XX

CC XX

CC XX

CC XX

CC XX

CC XX

CC XX

CC XX

CC XX

CC XX

CC XX

CC XX

CC XX

CC XX

CC XX

CC XX

CC XX

CC XX

Sequence 192, Application US/08467006  
 GENERAL INFORMATION:  
 APPLICANT: Griffith, Irwin J.  
 APPLICANT: Pollock, Joanne  
 APPLICANT: Bond, Julian F.  
 APPLICANT: Garman, Richard D.  
 APPLICANT: Kuo, Mei-Chang  
 APPLICANT: Yeung, Siu-mei H.  
 APPLICANT: Brauer, Andrew  
 APPLICANT: Exley, Mark A.  
 APPLICANT: Powers, Steven P.  
 TITLE OF INVENTION: Allergenic Proteins And Peptides From  
 TITLE OF INVENTION: Japanese Cedar Pollen  
 NUMBER OF SEQUENCES: 261  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Immunologic Pharmaceutical Corporation, Inc.  
 STREET: 610 Lincoln St.  
 CITY: Waltham  
 STATE: MA  
 COUNTRY: USA  
 ZIP: 02154  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent In Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/467,006  
 FILING DATE: June 6, 1995  
 CLASSIFICATION: 424  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 08/350,225  
 FILING DATE: December 6, 1994  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Jane E. Remillard  
 REGISTRATION NUMBER: 38,872  
 REFERENCE/DOCKET NUMBER: 025.6 USD6 (IMI-028CPD6)  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (617) 227-7400  
 TELEFAX: (617) 227-5941  
 INFORMATION FOR SEQ ID NO: 192:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 24 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: peptide



CC FRAGMENT TYPE: Internal  
SQ SEQUENCE 24 AA: 2551 MW: 3329 CN:  
  
Query Match 100.0% Score 72: DB 10: Length 24:  
Best Local Similarity 100.0% Pred. No. 1.63e-01:  
Matches 12: Conservative 0: Mismatches 0: Indels 0: Gaps 0:  
  
DB 13 ISLKTSGKIAS 24  
IIIIIIIIIIII  
QY 1 ISLKTSGKIAS 12  
  
RESULT 4  
ID US-06-468-940-192 STANDARD: PRT: 24 AA:  
XX AC xxxxxx  
XX AC xxxxxx  
DT  
XX DE  
XX DE Sequence 192, Application US/08468940  
XX DE  
XX DE Sequence 192, Application US/08468940  
CC GENERAL INFORMATION:  
CC APPLICANT: Griffith, Irwin J.:  
CC APPLICANT: Pollock, Joanne:  
CC APPLICANT: Bond, Julian F.:  
CC APPLICANT: Garman, Richard D:  
CC APPLICANT: Kuo, Mei-Chang:  
CC APPLICANT: Yeung, Siu-mei H.:  
CC APPLICANT: Brauer, Andrew:  
CC APPLICANT: Exley, Mark A.:  
CC APPLICANT: Powers, Steven P.:  
CC TITLE OF INVENTION: Allergenic Proteins And Peptides From  
CC TITLE OF INVENTION: Japanese Cedar Pollen  
CC NUMBER OF SEQUENCES: 261  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESS: Immunologic Pharmaceutical Corporation, Inc.  
CC STREET: 610 Lincoln St  
CC CITY: Waltham  
CC STATE: MA  
CC COUNTRY: USA  
CC ZIP: 02154  
CC  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: Patent In Release #1.0, Version #1.25  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/468,940  
CC FILING DATE:  
CC CLASSIFICATION: 424  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: 08/350,225  
CC FILING DATE: December 6, 1994  
CC APPLICATION NUMBER: 08/226,248  
CC FILING DATE: April 8, 1994  
CC APPLICATION NUMBER: 07/938,990  
CC FILING DATE: September 1, 1992  
CC APPLICATION NUMBER: PCT/US93/00139  
CC FILING DATE: January 15, 1993  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Darlene A. Vanstone  
CC REGISTRATION NUMBER: 35,729  
CC REFERENCE/DOCKET NUMBER: 025.6 US (IM1-028CP2)  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (617) 466-6000  
CC TELEFAX: (617) 466-6040  
CC INFORMATION FOR SEQ ID NO: 192:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 24 amino acids  
CC TYPE: amino acid  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: peptide

CC FRAGMENT TYPE: Internal  
SQ SEQUENCE 24 AA: 2551 MW: 3329 CN:  
  
Query Match 100.0% Score 72: DB 10: Length 24:  
Best Local Similarity 100.0% Pred. No. 1.63e-01:  
Matches 12: Conservative 0: Mismatches 0: Indels 0: Gaps 0:  
  
DB 13 ISLKTSGKIAS 24  
IIIIIIIIIIII  
QY 1 ISLKTSGKIAS 12  
  
RESULT 5  
ID US-08-226-248A-192 STANDARD: PRT: 24 AA:  
XX AC xxxxxx  
XX AC xxxxxx  
DT  
XX DE  
XX DE Sequence 192, Application US/08226248A  
XX DE  
XX DE Sequence 192, Application US/08226248A  
CC GENERAL INFORMATION:  
CC APPLICANT: Griffith, Irwin J.:  
CC APPLICANT: Pollock, Joanne:  
CC APPLICANT: Bond, Julian F.:  
CC APPLICANT: Garman, Richard D:  
CC APPLICANT: Kuo, Mei-Chang:  
CC APPLICANT: Yeung, Siu-mei H.:  
CC APPLICANT: Brauer, Andrew:  
CC APPLICANT: Exley, Mark A.:  
CC APPLICANT: Powers, Steven P.:  
CC TITLE OF INVENTION: Allergenic Proteins And Peptides From  
CC TITLE OF INVENTION: Japanese Cedar Pollen  
CC NUMBER OF SEQUENCES: 201  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESS: Immunologic Pharmaceutical Corporation, Inc.  
CC STREET: 610 Lincoln St  
CC CITY: Waltham  
CC STATE: MA  
CC COUNTRY: USA  
CC ZIP: 02154  
CC  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: Patent In Release #1.0, Version #1.25  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/226,248A  
CC FILING DATE: April 8, 1994  
CC CLASSIFICATION: 424  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: 07/938,990  
CC FILING DATE: September 1, 1992  
CC APPLICATION NUMBER: PCT/US93/00139  
CC FILING DATE: January 15, 1993  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Darlene A. Vanstone  
CC REGISTRATION NUMBER: 35,729  
CC REFERENCE/DOCKET NUMBER: 025.5 US (IM1-028CP)  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (617) 466-6000  
CC TELEFAX: (617) 466-6040  
CC INFORMATION FOR SEQ ID NO: 192:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 24 amino acids  
CC TYPE: amino acid  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: peptide  
CC FRAGMENT TYPE: Internal  
SQ SEQUENCE 24 AA: 2551 MW: 3329 CN:  
  
Query Match 100.0% Score 72: DB 7: Length 24:

```

Best Local Similarity 100.0%;      Pred. No. 1.63e-01;
Matches 12; Conservative 0;      Mismatches 0; Indels 0; Gaps 0;

```

```
DB 13 ISCKTSCX:AS 24  
      1 1111111 ,  
QY 1 ISCKTSCX:AS 24
```

RESULT 6  
ID US-68-350-225-102 STANDARD; PRT: 24 AA.  
XX  
AC XXXXX

RESULT: 7  
ID US-08-467-023-192 STANDARD: PRT: 24 AA.  
XX  
AC xxxxxx

|    |                                 |     |
|----|---------------------------------|-----|
| XX | Sequence 192, Application US/U: | 023 |
| DE |                                 |     |
| XX |                                 |     |
| CC | Sequence 192, Application US/   | 792 |

```

Db      13 ISLKLTSKIAS 24
        1111111111
QY      1 ISLKLTSKIAS 12

```



CC CURRENT APPLICATION DATA:  
 CC APPLICATION NUMBER: US/08/350.225  
 CC FILING DATE: December 6, 1994  
 CC CLASSIFICATION: 424  
 CC PRIOR APPLICATION DATA:  
 CC APPLICATION NUMBER: 08/226.248  
 CC FILING DATE: April 8, 1994  
 CC APPLICATION NUMBER: 07/938.990  
 CC FILING DATE: September 1, 1992  
 CC APPLICATION NUMBER: PC/US93/00139  
 CC FILING DATE: January 15, 1993  
 CC ATTORNEY/AGENT INFORMATION:  
 CC NAME: Carlene A. Vanstone  
 CC REGISTRATION NUMBER: 35,729  
 CC REFERENCE/DOCKET NUMBER: 025.6 US (.MI-028CP2)  
 CC TELECOMMUNICATION INFORMATION:  
 CC TELEPHONE: (517) 466-6090  
 CC TELEFAX: (517) 466-6040  
 CC INFORMATION FOR SEQ ID NO: 190:  
 CC SEQUENCE CHARACTERISTICS:  
 CC LENGTH: 127 amino acids  
 CC TYPE: amino acid  
 CC TOPOLOGY: linear  
 CC MOLECULE TYPE: peptide  
 CC FRAGMENT TYPE: internal  
 CC SEQUENCE 127 AA: 14435 MW: 84298 CN:

Query Match 100.0% Score 72: DB 8: Length 127:  
 Best Local Similarity 100.0% Pred. No. 1.63e-01:  
 Matches 12: Conservative 0: Mismatches 0: Indels 0:

DB 11 ISLKL7SGKIAS 22  
 QY 1 ISLKL7SGKIAS 12

RESULT 11  
 ID US-08-467-023-190 STANDARD: PRT: 127 AA.  
 XX xxxxxx  
 DE Sequence 190: Application US/08467023  
 CC Sequence 190: Application US/08467023  
 CC GENERAL INFORMATION:  
 CC APPLICANT: Griffith, Irwin J.  
 CC APPLICANT: Pollock, Joanne  
 CC APPLICANT: Bond, Julian F.  
 CC APPLICANT: Gorman, Richard D.  
 CC APPLICANT: Kuo, Mei-Chang  
 CC APPLICANT: Yeung, Siu-mei H.  
 CC APPLICANT: Brauer, Andrew  
 CC APPLICANT: Exley, Mark A.  
 CC APPLICANT: Powers, Steven P.  
 CC TITLE OF INVENTION: Allergenic Proteins And Peptides From  
 CC TITLE OF INVENTION: Japanese Cedar Pollen  
 CC NUMBER OF SEQUENCES: 261  
 CC CORRESPONDENCE ADDRESS:  
 CC ADDRESSEE: Immunologic Pharmaceutical Corporation, Inc.  
 CC STREET: 610 Lincoln St  
 CC CITY: Waltham  
 CC STATE: MA  
 CC COUNTRY: USA  
 CC ZIP: 02154  
 CC COMPUTER READABLE FORM:  
 CC MEDIUM TYPE: Floppy disk  
 CC COMPUTER: IBM PC compatible  
 CC OPERATING SYSTEM: PC-DOS/MS-DOS  
 CC SOFTWARE: Patent Release #1.0, Version #1.25  
 CC CURRENT APPLICATION DATA:  
 CC APPLICATION NUMBER: US/08/467.023

CC FILING DATE: June 6, 1995  
 CC CLASSIFICATION: 424  
 CC PRIOR APPLICATION DATA:  
 CC APPLICATION NUMBER: 08/350.225  
 CC FILING DATE: December 6, 1994  
 CC ATTORNEY/AGENT INFORMATION:  
 CC NAME: Jane E. Remillard  
 CC REGISTRATION NUMBER: 38,872  
 CC REFERENCE/DOCKET NUMBER: 025.6 US2 (.MI-028CP2)  
 CC TELECOMMUNICATION INFORMATION:  
 CC TELEPHONE: (617) 227-7400  
 CC TELEFAX: (617) 227-5941  
 CC INFORMATION FOR SEQ ID NO: 190:  
 CC SEQUENCE CHARACTERISTICS:  
 CC LENGTH: 127 amino acids  
 CC TYPE: amino acid  
 CC TOPOLOGY: linear  
 CC MOLECULE TYPE: peptide  
 CC FRAGMENT TYPE: internal  
 CC SEQUENCE 127 AA: 14435 MW: 84298 CN:  
 CC Query Match 100.0% Score 72: DB 10: Length 127:  
 CC Best Local Similarity 100.0% Pred. No. 1.63e-01:  
 CC Matches 12: Conservative 0: Mismatches 0: Indels 0:  
 CC DB 11 ISLKL7SGKIAS 22  
 CC QY 1 ISLKL7SGKIAS 12  
 CC RESULT 12  
 CC ID US-08-226-248A-190 STANDARD: PRT: 127 AA.  
 CC XX xxxxxx  
 CC DE Sequence 190: Application US/08226248A  
 CC Sequence 190: Application US/08226248A  
 CC GENERAL INFORMATION:  
 CC APPLICANT: Griffith, Irwin J.  
 CC APPLICANT: Pollock, Joanne  
 CC APPLICANT: Bond, Julian F.  
 CC APPLICANT: Gorman, Richard D.  
 CC APPLICANT: Kuo, Mei-Chang  
 CC APPLICANT: Yeung, Siu-mei H.  
 CC APPLICANT: Brauer, Andrew  
 CC APPLICANT: Exley, Mark A.  
 CC APPLICANT: Powers, Steven P.  
 CC TITLE OF INVENTION: Allergenic Proteins And Peptides From  
 CC TITLE OF INVENTION: Japanese Cedar Pollen  
 CC NUMBER OF SEQUENCES: 201  
 CC CORRESPONDENCE ADDRESS:  
 CC ADDRESSEE: Immunologic Pharmaceutical Corporation, Inc.  
 CC STREET: 610 Lincoln St  
 CC CITY: Waltham  
 CC STATE: MA  
 CC COUNTRY: USA  
 CC ZIP: 02154  
 CC COMPUTER READABLE FORM:  
 CC MEDIUM TYPE: Floppy disk  
 CC COMPUTER: IBM PC compatible  
 CC OPERATING SYSTEM: PC-DOS/MS-DOS  
 CC SOFTWARE: Patent Release #1.0, Version #1.25  
 CC CURRENT APPLICATION DATA:  
 CC APPLICATION NUMBER: US/08/226.248A  
 CC FILING DATE: April 8, 1994  
 CC CLASSIFICATION: 424  
 CC PRIOR APPLICATION DATA:  
 CC APPLICATION NUMBER: 07/938.990  
 CC FILING DATE: September 1, 1992  
 CC APPLICATION NUMBER: PC/US93/00139

CC FILING DATE: January 15, 1993  
 CC ATTORNEY/AGENT INFORMATION:  
 CC NAME: Darlene A. Vanstone  
 CC REGISTRATION NUMBER: 35,729  
 CC REFERENCE/DOCKET NUMBER: 025.5 US (IMI-028CP)  
 CC TELECOMMUNICATION INFORMATION:  
 CC TELEPHONE: (617) 466-6000  
 CC TELEFAX: (617) 466-6040  
 CC INFORMATION FOR SEQ ID NO: 190:  
 CC SEQUENCE CHARACTERISTICS:  
 CC LENGTH: 127 amino acids  
 CC TYPE: amino acid  
 CC TOPOLOGY: linear  
 CC MOLECULE TYPE: peptide  
 CC FRAGMENT TYPE: internal  
 CC SEQUENCE 127 AA: 14435 MW: 84298 CN:  
 SC

Query Match 100.0% Score 72: DB 7: Length 127:  
 Best Local Similarity 100.0% Pred. No. 1,53e-01:  
 Matches 12: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

DB 11 ISLKLTSKGIAS 22  
 QY 1 ISLKLTSKGIAS 12  
 RESULT 13  
 ID US-08-467-697-190 STANDARD: PRT: 127 AA.  
 XX xxxxxx  
 XX DT  
 XX DE  
 XX SEQUENCE 190. Application US/08457697  
 CC Sequence 190. Application US/08457697  
 CC GENERAL INFORMATION:  
 CC APPLICANT: Griffith, Irwin J.  
 CC APPLICANT: Pollock, Joanne  
 CC APPLICANT: Bond, Julian F.  
 CC APPLICANT: Garman, Richard D.  
 CC APPLICANT: Kuo, Mei-Chang H.  
 CC APPLICANT: Yeung, Siu-mei H.  
 CC APPLICANT: Brauer, Andrew  
 CC APPLICANT: Exley, Mark A.  
 CC APPLICANT: Powers, Steven P.  
 CC TITLE OF INVENTION: Allergenic Proteins And Peptides From  
 CC TITLE OF INVENTION: Japanese Cedar Pollen  
 CC NUMBER OF SEQUENCES: 251  
 CC CORRESPONDENCE ADDRESS:  
 CC ADDRESS: Immunologic Pharmaceutical Corporation, Inc.  
 CC STREET: 610 Lincoln St  
 CC CITY: Waltham  
 CC STATE: MA  
 CC COUNTRY: USA  
 CC ZIP: 02154  
 CC COMPUTER READABLE FORM:  
 CC MEDIUM TYPE: Floppy disk  
 CC COMPUTER: IBM PC compatible  
 CC OPERATING SYSTEM: PC-DOS/MS-DOS  
 CC SOFTWARE: PatentIn Release #1.0, Version #1.25  
 CC CURRENT APPLICATION DATA:  
 CC APPLICATION NUMBER: US/08/467,697  
 CC FILING DATE: June 6, 1995  
 CC CLASSIFICATION: 424  
 CC PRIOR APPLICATION DATA:  
 CC APPLICATION NUMBER: 08/350,225  
 CC FILING DATE: December 6, 1994  
 CC ATTORNEY/AGENT INFORMATION:  
 CC NAME: Jane E. Remillard  
 CC REGISTRATION NUMBER: 38,872  
 CC REFERENCE/DOCKET NUMBER: 025.6 USD4 (IMI-028CPD4)  
 CC TELECOMMUNICATION INFORMATION:

CC TELEPHONE: (617) 227-7400  
 CC TELEFAX: (617) 227-5941  
 CC INFORMATION FOR SEQ ID NO: 190:  
 CC SEQUENCE CHARACTERISTICS:  
 CC LENGTH: 127 amino acids  
 CC TYPE: amino acid  
 CC TOPOLOGY: linear  
 CC MOLECULE TYPE: peptide  
 CC FRAGMENT TYPE: internal  
 CC SEQUENCE 127 AA: 14435 MW: 84298 CN:  
 SC

Query Match 100.0% Score 72: DB 10: Length 127:  
 Best Local Similarity 100.0% Pred. No. 1,53e-01:  
 Matches 12: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

DB 11 ISLKLTSKGIAS 22  
 QY 1 ISLKLTSKGIAS 12  
 RESULT 14  
 ID US-08-468-940-190 STANDARD: PRT: 127 AA.  
 XX xxxxxx  
 XX DT  
 XX DE  
 XX SEQUENCE 190. Application US/08458940  
 CC Sequence 190. Application US/08458940  
 CC GENERAL INFORMATION:  
 CC APPLICANT: Griffith, Irwin J.  
 CC APPLICANT: Pollock, Joanne  
 CC APPLICANT: Bond, Julian F.  
 CC APPLICANT: Garman, Richard D.  
 CC APPLICANT: Kuo, Mei-Chang H.  
 CC APPLICANT: Yeung, Siu-mei H.  
 CC APPLICANT: Brauer, Andrew  
 CC APPLICANT: Exley, Mark A.  
 CC APPLICANT: Powers, Steven P.  
 CC TITLE OF INVENTION: Allergenic Proteins And Peptides From  
 CC TITLE OF INVENTION: Japanese Cedar Pollen  
 CC NUMBER OF SEQUENCES: 251  
 CC CORRESPONDENCE ADDRESS:  
 CC ADDRESS: Immunologic Pharmaceutical Corporation, Inc.  
 CC STREET: 610 Lincoln St  
 CC CITY: Waltham  
 CC STATE: MA  
 CC COUNTRY: USA  
 CC ZIP: 02154  
 CC COMPUTER READABLE FORM:  
 CC MEDIUM TYPE: Floppy disk  
 CC COMPUTER: IBM PC compatible  
 CC OPERATING SYSTEM: PC-DOS/MS-DOS  
 CC SOFTWARE: PatentIn Release #1.0, Version #1.25  
 CC CURRENT APPLICATION DATA:  
 CC APPLICATION NUMBER: US/08/468,940  
 CC FILING DATE:  
 CC CLASSIFICATION: 424  
 CC PRIOR APPLICATION DATA:  
 CC APPLICATION NUMBER: 08/350,225  
 CC FILING DATE: December 6, 1994  
 CC ATTORNEY/AGENT INFORMATION:  
 CC NAME: Darlene A. Vanstone  
 CC REGISTRATION NUMBER: 35,729  
 CC REFERENCE/DOCKET NUMBER: 025.6 US (IMI-028CP2)  
 CC TELECOMMUNICATION INFORMATION:

CC TELEPHONE: (617) 466-6000  
CC TELEFAX: (617) 466-6040  
CC INFORMATION FOR SEQ ID NO: 190:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 127 amino acids  
CC TYPE: amino acid  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: peptide  
CC FRAGMENT TYPE: internal  
CC SEQUENCE 127 AA: 14435 MW: 84258 CN:  
Query Match 100.0% Score 72: DB 10: Length 127:  
Best Local Similarity 100.0% Pred. No. 1.63e-01:  
Matches 12: Conservative 0: Mismatches 0: Indels 0: Gaps 0:  
DB 11 ISKLTSGKIAS 22  
QY 1 ISKLTSGKIAS 12  
RESULT 15  
ID US-08-571-978-1 STANDARD: PRT: 460 AA  
AC xxxxxx  
CT  
DE Sequence 1: Application US/08571978  
CC Sequence 1: Application US/08571978  
CC GENERAL INFORMATION:  
CC APPLICANT: KINO, KOSUKE  
CC APPLICANT: KOMIYAMA, NAOKI  
CC APPLICANT: SONE, TOSHIO  
CC APPLICANT: KOHNO, YOICHI  
CC TITLE OF INVENTION: ANTI-ALLERGIC AGENTS  
CC NUMBER OF SEQUENCES: 36  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,  
CC STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400  
CC CITY: ARLINGTON  
CC STATE: VA  
CC COUNTRY: USA  
CC ZIP: 22202  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: floppy disk  
CC COMPUTER: IBM PC Compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: PatentIn Release #1.0, Version #1.30  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/571,978  
CC FILING DATE: 15-JAN-1996  
CC CLASSIFICATION: 514  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: PCT/JP94/01164  
CC FILING DATE: 15-JUL-1994  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: JP 6-069336  
CC FILING DATE: 07-APR-1994  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: JP 5-217725  
CC FILING DATE: 01-SEP-1993  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: JP 5-177008  
CC FILING DATE: 16-JUL-1993  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: OHLON, NORMAN F.  
CC REGISTRATION NUMBER: 24,618  
CC REFERENCE/CKET NUMBER: 7218-002-0 PCT  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: 703-413-3000  
CC TELEFAX: 703-413-2220

CC INFORMATION FOR SEQ ID NO: 1:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 460 amino acids  
CC TYPE: amino acid  
CC STRANDEDNESS: single  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: peptide  
CC SEQUENCE 460 AA: 50440 MW: 1058702 CN:  
Query Match 100.0% Score 72: DB 11: Length 460:  
Best Local Similarity 100.0% Pred. No. 1.63e-01:  
Matches 12: Conservative 0: Mismatches 0: Indels 0: Gaps 0:  
DB 344 ISKLTSGKIAS 355  
QY 1 ISKLTSGKIAS 12  
Search completed: Mon Jun 19 16:24:10 2000  
Job time : 20 secs.

\*\*\*\*\*  
 RELEASE  
 \*\*\*\*\*  
 (TM)  
 \*\*\*\*\*  
 Release 3.1A John F. Collins, BioComputing Research Unit.  
 Copyright (c) 1993-1998 University of Edinburgh, U.K.  
 Distribution rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm  
 Run on: Mon Jun 19 16:04:40 2000: MasPar time 3.89 Seconds  
 145.422 Million cell updates/sec  
 Tabular output not generated.

Title: >US-09-142-524A-6  
 Description: {1-12} from US09142524A.ppt  
 Perfect Score: 72  
 Sequence: 1 ISLKLTSKGIAS 12

Scoring table: PAM 150  
 Gap 15

Searched: 142080 seqs, 47172406 residues

Post-processing: Minimum Match 08

Listing first 45 summaries

Database: Pf62  
 1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 23.729; Variance 26.562; scale 0.893

Pred. No. is the number of results predicted by chance to have a  
 score greater than or equal to the score of the result being printed,  
 and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID     | Description             | Pred. No. |
|------------|-------|-------------|--------|----|--------|-------------------------|-----------|
| 1          | 72    | 100.0       | 514    | 2  | JC2498 | second major allergen   | 1.81e-05  |
| 2          | 72    | 100.0       | 514    | 2  | S48730 | Cry j II protein - Ja   | 1.81e-05  |
| 3          | 52    | 72.2        | 182    | 2  | F7C509 | hypothetical protein    | 1.73e+00  |
| 4          | 51    | 70.8        | 255    | 1  | ORECM3 | membrane-bound iron(I   | 2.90e+00  |
| 5          | 49    | 68.1        | 325    | 2  | I04273 | hypothetical protein    | 7.94e+00  |
| 6          | 49    | 68.1        | 407    | 2  | E69409 | conserved hypotheticala | 7.94e+00  |
| 7          | 49    | 68.1        | 412    | 1  | G70017 | N-carbamyl-L-amino ac   | 7.94e+00  |
| 8          | 49    | 68.1        | 573    | 1  | S41375 | leucyl aminopeptidase   | 7.94e+00  |
| 9          | 48    | 66.7        | 210    | 2  | T05298 | hypothetical protein    | 1.30e+01  |
| 10         | 48    | 66.7        | 309    | 2  | A70194 | ABC transporter, ATP-   | 1.30e+01  |
| 11         | 48    | 66.7        | 376    | 2  | S64479 | probable membrane pro   | 1.30e+01  |
| 12         | 48    | 66.7        | 460    | 1  | E70132 | hypothetical protein    | 1.30e+01  |
| 13         | 48    | 66.7        | 1019   | 1  | A45341 | major structural core   | 1.30e+01  |
| 14         | 48    | 66.7        | 1120   | 2  | H71604 | transcription-repair    | 1.30e+01  |
| 15         | 48    | 66.7        | 1585   | 2  | BG9948 | phage-related protein   | 1.30e+01  |
| 16         | 48    | 66.7        | 1743   | 2  | T15893 | hypothetical protein    | 1.30e+01  |
| 17         | 47    | 65.3        | 155    | 2  | F75125 | hypothetical protein    | 2.11e+01  |
| 18         | 47    | 65.3        | 199    | 2  | H69439 | conserved hypotheticala | 2.11e+01  |
| 19         | 47    | 65.3        | 298    | 2  | S77536 | lactose transport sys   | 2.11e+01  |
| 20         | 47    | 65.3        | 530    | 1  | W8BE04 | gene 40 protein - equ   | 2.11e+01  |
| 21         | 47    | 65.3        | 572    | 2  | S68119 | laccase (EC 1.10.3.2)   | 2.11e+01  |
| 22         | 47    | 65.3        | 694    | 2  | G72009 | polyribonucleotide nu   | 2.11e+01  |
| 23         | 47    | 65.3        | 907    | 2  | J60193 | G protein-coupled rec   | 2.11e+01  |

| 24 | 46 | 63.9 | 121  | 2 | A71    | 1 | hypothetical protein   | 3.40e-01 |
|----|----|------|------|---|--------|---|------------------------|----------|
| 25 | 46 | 63.9 | 124  | 1 | G61    | 1 | hydrogenase accessory  | 3.40e-01 |
| 26 | 46 | 63.9 | 145  | 2 | C55824 | 1 | agbB protein precursor | 3.40e-01 |
| 27 | 46 | 63.9 | 251  | 2 | B64948 | 1 | probable ABC transpor  | 3.40e-01 |
| 28 | 46 | 63.9 | 286  | 2 | S64311 | 1 | beta-lactamase (EC 3.  | 3.40e-01 |
| 29 | 46 | 63.9 | 286  | 4 | I40945 | 1 | beta-lactamase (EC 3.  | 3.40e-01 |
| 30 | 46 | 63.9 | 286  | 4 | S41475 | 1 | beta-lactamase (EC 3.  | 3.40e-01 |
| 31 | 46 | 63.9 | 286  | 2 | S30113 | 1 | beta-lactamase (EC 3.  | 3.40e-01 |
| 32 | 46 | 63.9 | 286  | 1 | PNE1P  | 1 | beta-lactamase (EC 3.  | 3.40e-01 |
| 33 | 46 | 63.9 | 286  | 2 | S41116 | 1 | Bla protein - Salmone  | 3.40e-01 |
| 34 | 46 | 63.9 | 286  | 2 | S60320 | 1 | extended spectrum bet  | 3.40e-01 |
| 35 | 46 | 63.9 | 286  | 2 | S61110 | 1 | beta-lactamase (EC 3.  | 3.40e-01 |
| 36 | 46 | 63.9 | 317  | 2 | S41110 | 1 | transaldolase (EC 2.2  | 3.40e-01 |
| 37 | 46 | 63.9 | 359  | 2 | S54110 | 1 | alpha-N-acetyltransfer | 3.40e-01 |
| 38 | 46 | 63.9 | 359  | 2 | S54110 | 1 | alpha-2,8-polysialylt  | 3.40e-01 |
| 39 | 46 | 63.9 | 359  | 2 | S54110 | 1 | polysialyltransferase  | 3.40e-01 |
| 40 | 46 | 63.9 | 359  | 2 | S54110 | 1 | N3L protein - vaccini  | 3.40e-01 |
| 41 | 46 | 63.9 | 551  | 1 | Q4V221 | 1 | N3L protein - vaccini  | 3.40e-01 |
| 42 | 46 | 63.9 | 551  | 2 | A36848 | 1 | hemoglobin receptor h  | 3.40e-01 |
| 43 | 46 | 63.9 | 758  | 2 | I64084 | 1 | probable DNA-directed  | 3.40e-01 |
| 44 | 46 | 63.9 | 1202 | 1 | S05362 | 1 | polyketide synthase p  | 3.40e-01 |
| 45 | 46 | 63.9 | 4273 | 2 | C69679 | 1 |                        |          |

#### ALIGNMENTS

RESULT 1  
 ENTRY #type complete  
 TITLE second major allergen Cry j II precursor - Japanese cedar  
 ORGANISM #formal name Cryptomeria japonica #common name Japanese cedar  
 DATE 16-Mar-1995 #sequence\_revision 26-May-1995 #text\_change

12-Sep-1997  
 JC2498: PC2346; A60147  
 JC2498  
 Komiya, N.; Sone, I.; Shimizu, K.; Morikubo, K.; Kuno, K.  
 Biochem Biophys Res Commun (1994) 201:1021-1029  
 cDNA cloning and expression of Cry j II, the second major  
 allergen of Japanese cedar pollen.

#cross-references MGID:94271186  
 #accession JC2498  
 #molecule\_type mRNA  
 #residues 1-514 #label: KOM  
 #cross-references DDBJ:D297 J NID:9506857; PID:dl006744; PIR:U556458  
 #accession PC2346  
 #molecule\_type protein  
 #residues 52-61 #label: 492  
 #reference A60147  
 Sakaguchi, M.; Imoye, S.; Tabata, M.; Ando, S.; Usui, M.;  
 Matsumoto, T.

#journal Allergy (1990) 45:309-312  
 #title Identification of the second major allergen of Japanese cedar  
 #cross-references MGID:90342988  
 #accession A60147  
 #molecule\_type protein  
 #residues 55-64 #label: SAK  
 #keywords glycoprotein; pollen  
 #feature 1-54  
 #domain signal sequence #status predicted #label SIG  
 #product second major allergen Cry j #status predicted  
 #label MATN  
 #binding-site carbohydrate (Asn) (covalent) #status  
 predicted  
 429,450,472

SUMMARY  
 #length 514 #mole 1 #ar-weight 56583 #checksum 975  
 Query Match 100.0%; Score 72; DB 2; Length 514;  
 Best Local Similarity 100.0%; Pred. No. 1.81e-05;  
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 398 ISLKLTSKGIAS 409  
 OY 1 ISLKLTSKGIAS 12

```

QY 398 ISLKTSGKIA 12
RESULT 2
ENTRY Cry J II protein - Japanese cedar
TITLE #formal_name Cryptomeria japonica #common_name Japanese cedar
ORGANISM 07-May-1995 #sequence_revision 21-Jul-1995 #text_change
DATE 28-May-1999
ACCESSIONS S48730
REFERENCE Narba, M.; Kurose, M.; Torigoe, K.; Hino, K.; Taniguchi, Y.; Fukuda, S.; Usui, M.; Kurimoto, M. FEBS Lett. (1994) 353:124-128. #journal Molecular cloning of the second major allergen, Cry J II, from Japanese cedar pollen. #title #cross-references MIM:601077 #accession S48730 #status preliminary #molecule_type mRNA #residues 1-514 #label NAM #cross-references GB:D37765; NID:q577695; PIDN:BAAU7021.1; PID:q1007598; PID:q577696 #length 514 #molecular_weight 56645 #checksum 585
SUMMARY Query Match 100.0%; Score 72; DB 2; Length 514; Best Local Similarity 100.0%; Pred. No. 1:8e-05; Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ISLKTSGKIA 12
RESULT 4
ENTRY QREOM3
TITLE membrane-bound iron(III) citrate transport protein - Escherichia coli
ORGANISM #formal_name Escherichia coli
DATE 31-Dec-1989 #sequence_revision 31-Dec-1989 #text_change
ACCESSIONS J50135; S56512; A65242
REFERENCE Staudenmaier, H.; Van Hove, B.; Varaghi, Z.; Braun, V. J. Bacteriol. (1989) 171:2646-2653. #journal Nucleotide sequences of the fecABCDE genes and locations of the proteins suggest a periplasmic-binding protein-dependent transport mechanism for iron(III) citrate in Escherichia coli. #title #cross-references MIM:89213950 #accession J50135 #molecule_type DNA #residues 1-255 #label STA #cross-references GB:M26397; NID:q145928; PIDN:AAA2765.1; PID:q145928 #experimental_source strain #12
REFERENCE S56512
#authors Blattner, F.R.; Pluckert, J.L.; Sofia, H.E.; Daniels, L.L.; Nucleic Acids Res. (1995) 23:2105-2119. #journal Analysis of the Escherichia coli genome VI: DNA sequence of the region from 9.8 through 10.6 minutes. #title #cross-references MIM:95334362 #accession S56512 #status preliminary; nucleic acid sequence not shown; translation not shown
#molecule_type DNA #residues 1-255 #label BJR #cross-references EMBL:014303; NID:q1263172; PIDN:AAA97183.1; PID:q537128 #note the nucleotide sequence was submitted to the EMBL Data Library, August 1994
REFERENCE A64720
#authors Blattner, F.R.; Pluckert, J.L.; Sofia, H.E.; Perna, N.T.; Blattner, V.; Riley, M.; Collado-Vides, J.; Glasner, J.E.; Rode, C.K.; Mayhew, G.F.; Greider, J.; Davis, N.W.; Kirkpatrick, H.A.; Peterson, M.A.; Rose, D.J.; Mau, H.; Shao, Y. Science (1997) 275:109-115. #journal The complete genome sequence of Escherichia coli K12. #title #cross-references MIM:97426617 #accession A65242 #status preliminary; nucleic acid sequence not shown; translation not shown
#molecule_type DNA #residues 1-255 #label BLAT #cross-references GB:AE000499; GB:U00096; NID:q1790732; PIDN:AA07243.1; PID:q1790739; UWGP:B4287 #experimental_source strain K12, substrain MG1655
GENETICS
#gene fecE
#map_position 7 min
#description one of five, encoded by the fec operon, constituting a citrate-dependent iron(III) transport system
CLASSIFICATION #superfamily inner membrane protein malk; ATP-binding cassette homology
KEYWORDS ATP; iron transport; membrane protein; nucleotide binding; P-loop
FEATURE 18-214
35-42
SUMMARY
Query Match 70.8%; Score 51; DB 1; Length 255;

```



```

Best Local Similarity 50.08: Pred. No. 2,9: 3-00;
Matches 6: Conservative 4: Mismatches 2: Indels 0: Gaps 0;

Db 21 VS-SLPTGKITA 32
    ||| |||
QY 1 ISLKLTSCKIA 12

RESULT 5
ENTRY T04273 #type complete
TITLE hypothetical protein F20B18.290 - Arabidopsis thaliana
ORGANISM #formal_name Arabidopsis thaliana #common_name mouse-ear
        cress
DATE 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change
        20-Sep-1999
ACCESSIONS 204273
REFERENCE 219263
AUTHORS Bryan, M.; Rose, M.; Hempel, S.; Entian, K.D.; Hobetsel, J.;
        Newes, H.W.; Mayer, K.F.X.; Schueller, C.
#submission submitted to the Protein Sequence Database, March 1999
#accession T04273
#molecule_type DNA
#residues 1-325 #label BEV
#cross-references EMBL:AF049483
#experimental_source cultivar Columbia; BAC clone F20B18
GENETICS
#map_position 4
#introns F20B18.290
#note
CLASSIFICATION
#superfamily ADP,ATP carrier protein; ADP,ATP carrier protein
#repeat_homology
#length 325 #molecular_weight 36551 #checksum 6140
SUMMARY
Query Match 58.18: Score 49: DB 2: Length 325:
Best Local Similarity 58.38: Pred. No. 7,948-00;
Matches 7: Conservative 2: Mismatches 3: Indels 0: Gaps 0;

Db 213 ISLKLVCSSVAG 224
    |||||
QY 1 ISLKLTSCKIA 12

RESULT 6
ENTRY T04273 #type complete
TITLE conserved hypothetical protein AF1278 - Archaeoglobus
        fulgidus
ORGANISM #formal_name Archaeoglobus fulgidus
        05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change
        05-Jun-1998
ACCESSIONS F09409
REFERENCE AK9250
AUTHORS Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson,
        K.E.; Ketchum, K.A.; Dodson, R.J.; Quinn, M.; Hickey, E.K.;
        Peterson, J.D.; Richardson, D.L.; Kurlavake, A.R.; Graham,
        D.E.; Kyprides, N.C.; Fleischmann, R.D.; Quackenbush, J.;
        Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.;
        Dougherty, S.A.; McKenry, K.; Adams, M.D.; Loftis, B.;
        Petersen, S.; Reich, C.I.; McNeil, L.K.; Badger, J.H.;
        Glöck, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman,
        J.F.; McDonald, L.; Utterback, T.; Cotton, M.D.; Spriggs,
        T.; Artiach, P.; Kaine, B.P.; Sykes, S.M.; Sadow, P.W.;
        D'Andrea, K.P.; Bowman, C.; Fujii, C.; Garland, S.A.;
        Mason, T.M.; Olsen, G.J.; Fraser, C.M.; Smith, H.O.; Woese,
        C.R.; Venter, J.C.
#journal Nature (1997) 390:364-370
#title The complete genome sequence of the hyperthermophilic,
        sulfate-reducing archaeon Archaeoglobus fulgidus.
#cross-references M31D:98049343
#accession F09409
#status preliminary: nucleic acid sequence not shown;
        translation not shown
#molecule_type DNA
#residues 1-407 #label KLE

```

```

**cross-references GB:AE001016: GB:AE000782: NID:q2589339: PID:q2649302:
TIGR:AF1278
SUMMARY #length 407 #molecular_weight 44931 #checksum 14
Query Match 68.18: Score 49: DB 2: Length 407:
Best Local Similarity 77.88: Pred. No. 7,948-00;
Matches 7: Conservative 1: Mismatches 1: Indels 0: Gaps 0;

Db 193 LKLKCKIA 201
    ||| ||||
QY 3 LKLTSCKIA 11

RESULT 7
ENTRY G70017 #type complete
TITLE N-carbamyl-L-amino acid amidohydrolase homolog yurH -
        Bacillus subtilis
ORGANISM #formal_name Bacillus subtilis
DATE 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change
        10-Sep-1999
ACCESSIONS G70017
REFERENCE A69580
AUTHORS Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.;
        Altoni, G.; Azevedo, V.; Bertero, M.G.; Bessières, P.;
        Biotin, A.; Borchert, S.; Boriss, R.; Borriss, J.; Brans,
        A.; Braun, M.; Brignell, S.C.; Bron, S.; Broillet, S.;
        Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
        Choi, S.K.; Codanis, J.J.; Connerton, I.F.; Cummings, N.J.;
        Daniel, R.A.; Denitot, F.; Devine, K.M.; Duersthoefel, A.;
        Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Enting, J.;
        Fabret, C.; Ferrari, E.; Foulger, D.; Fritz, C.; Fujita,
        M.; Fujita, Y.; Fuma, S.; Gallizi, A.; Gallier, N.; Ghim,
        S.Y.; Glaser, P.; Goffeau, A.; Golightly, E.J.; Grandi, G.;
        Giuseppe, G.; Guy, B.J.; Haga, K.; Halech, J.; Harwood,
        C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Husono, S.;
        Hullo, M.F.; Itaya, M.; Jones, L.; Joris, S.; Karamata, D.;
        Kasahara, Y.; Klier-Blanchard, M.; Klein, C.; Kobayashi,
        Y.; Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.;
        Kurita, K.; Lapidus, A.; Lardinois, S.; Lamber, J.;
        Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.;
        Maueel, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno,
        M.; Moestl, D.; Nakai, S.; Noback, M.; Noone, E.; O'Reilly,
        M.; Ogawa, K.; Ogihara, A.; Oudega, B.; Park, S.H.; Parré,
        V.; Pohl, T.M.; Portetle, D.; Porwollik, S.; Prestcott,
        A.M.; Prosser, E.; Pulic, P.; Purnelle, S.; Ravipati, S.;
        Key, M.; Reynolds, S.; Rinder, M.; Rivalland, J.; Roth, E.;
        Kocay, B.; Rose, M.; Saday, Y.; Sato, T.; Schmitt, E.;
        Schreier, S.; Schreiber, K.; Scellone, F.; Sekiguchi, J.;
        Sekowska, A.; Serot, S.; Serrón, P.; Shin, B.S.; Soldo,
        B.; Sorokin, A.; Tacconi, E.; Takaki, T.; Takahashi, H.;
        Takemaru, K.; Takeuchi, M.; Tamakoshi, A.; Tanaka, T.;
        Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.;
        Vandenbol, M.; Vannier, F.; Vassarotti, A.; Viñari, A.;
        Wambutt, R.; Wedler, E.; Wedler, H.; Weitzengraber, T.;
        Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasunoto,
        K.; Yata, K.; Yoshida, K.; Yoshikawa, H.F.; Zimstein, E.;
        Yoshikawa, H.; Zechin, A.
#journal Nature (1997) 390:249-256
#title The complete genome sequence of the Gram-positive bacterium
        Bacillus subtilis s.
#cross-references M31D:98044033
#accession G70017
#status preliminary: nucleic acid sequence not shown;
        translation not shown
#molecule_type DNA
#residues 1-412 #label KUN
#cross-references GB:Z99120: GB:AL009126: NID:q2635613:
        PID:q2635750
        PID:q2635750
#experimental_source strain 168
GENETICS
#gene yurH
CLASSIFICATION
#superfamily N-carbamyl-L-amino acid amidohydrolase
#length 412 #molecular_weight 45519 #checksum 7655
SUMMARY

```

```

##experimental_source cultivar Columbia: BAC clone F26p21

GENETICS
  #map_position 4
  #introns      14/3: 8/3: 102/1: 114/2
  #note         F26p21.20
  #length 210  #molecular_weight 23517  #checksum 224

Query Match      66.7%  Score 48:  DB 2: Length 210:
Best Local Similarity 70.0%:  Pred. NO. 1,30e-01:
Matches      7:  Conservative      2:  Mismatches      1:  Indels      0:  Gaps      0:

Db      89 LKMSGKIAS 58
QY      3 LKLSGKIAS 12
      11: 1111
      11: 1111

RESULT      10
ENTRY      A70134  #type complete
TITLE      ABC transporter, ATP-binding protein homologs - Tyne disease
           #spirochete
ORGANISM   Borrelia burgdorferi #common_name Tyne disease
           #spirochete
DATE       13-Feb-1998 #seq_rev 13-Feb-1998 #text_change
           28-May-1999
ACCESSIONS A70134
REFERENCE  A70194
#authors   Fraser, C.M.; Castens, S.; Huang, M.M.; Ketchum, C.A.;
           Dodson, R.; Ladhari, K.; White, O.; Ketchum, K.A.; Tish,
           J.F.; Fleischman, R.D.; Richardson, D.; Peterson, J.;
           Kevlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.;
           Vost, R.V.; Palmer, N.; Adams, M.D.; Gocayne, J.; Weidman,
           J.; Utterback, T.; Matthey, L.; McDonald, L.; Artach, P.;
           Bowman, C.; Garland, S.; Fujita, C.; Cotton, M.; Hest,
           K.; Roberts, K.; Hatch, R.; Smith, H.G.; Venter, J.;
           Nature (1997) 390:580-582
#title     Genomic sequence of a Tyne disease spirochaete, Borrelia
           burgdorferi.
#cross-references MIM:19806594;
#accession A70194
#status    preliminary: nucleic acid sequence not shown;
           translation not shown
#molecule_type DNA
#residues   1-309 #label K1F
#cross-references GB:AE001101; GB:AB007943; NC_024665; EMBL:U88689;
           F53897.4;
           #spirochete
#experimental_source strain B41
#classification superfamily ATP-binding cassette homology
KEYWORDS    ATP; p-loop
FEATURE     17-205
           34-41
           #domain, ATP-binding cassette homology #label ABC
           #region nucleotide-binding motif A (p-loop)
SUMMARY     #length 309  #molecular_weight 35183  #checksum 7272

Query Match      66.7%  Score 48:  DB 2: Length 309:
Best Local Similarity 77.8%:  Pred. NO. 1,30e-01:
Matches      7:  Conservative      1:  Mismatches      1:  Indels      0:  Gaps      0:

Db      262 ISLKSQK 270
QY      1 ISLKL-SGK 9
      1111: 11
      1111: 11

RESULT      1:
ENTRY      S64479  #type complete
TITLE      Probable membrane protein YGR169c - yeast (Saccharomyces
           cerevisiae)
ALTERNATE_NAMES hypothetical protein G7059
ORGANISM    Saccharomyces cerevisiae
#formal_name Saccharomyces cerevisiae
DATE       17-May-1996 #sequence_revision 17-May-1996 #text_change
           21-Nov-1997
ACCESSIONS S64479; S64482
REFERENCE  S64071

```

```

#authors Rieger, M.; Mueller-Auer, S.; Brueckner, M.; Schaefer, M.
#submission Submitted to the Protein Sequence Database, May 1996
#accession S64479
##molecule_type DNA
##residues 1-376 ##label RIE
##cross-references EMBL:272953; NID:g1323295; PID:e243553; PID:g1323296;
MIPS:YGR168C
##experimental_source strain S288C
REFERENCE
S54003
#authors Hobling, U.; Hofmann, B.; Deilus, H.
#submission Submitted to the Protein Sequence Database, May 1996
#accession S64482
##molecule_type DNA
##residues 1-376 ##label HEB
##cross-references EMBL:272953; NID:g1323295; PID:e243553; PID:g1323296;
MIPS:YGR168C
##experimental_source strain S288C
GENETICS
#map_position 78
#transmembrane protein
KEYWORDS
61-97
#domain transmembrane #status predicted #label TM1\
156-172
#domain transmembrane #status predicted #label TM2\
191-207
#domain transmembrane #status predicted #label TM3\
SUMMARY
#length 376 #molecular-weight 44301 #checksum 9227
Query Match 66.7%; Score 48; DB 2; Length 376;
Best Loca: Similarity 58.3%; Pred. No. 1.30e+01;
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
DB 111 ISLMLTKGEVPS 142
1111111111
QY 1 ISLKLTSCKIAS 12
RESULT 12
ENTRY
#type complete
#hypothetical protein BB0261 - Lyme disease spirochete
#formal_name Borrelia burgdorferi #common_name Lyme disease
#spirochete
#10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change
10-Sep-1999
ACCESSIONS
E70132
REFERENCE
#authors
Fraser, C.M.; Castjens, S.; Huang, W.M.; Sutton, G.G.;
Clayton, R.; Lathigra, R.; White, O.; Ketchum, K.A.;
Daxson, R.; Hickey, E.K.; Gwinn, M.; Dougherty, B.; Tomb,
J.F.; Fleischmann, R.D.; Richardson, D.; Peterson, J.;
Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.;
Vost, R.V.; Palmer, N.; Adams, M.D.; Gocayne, J.; Weidman,
J.; Utterback, T.; Watthey, L.; McDonald, L.; Artiach, P.;
Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst,
K.; Roberts, K.; Hatch, B.; Smith, R.C.; Venter, J.C.
#journal Nature (1997) 390:580-586
#title Genomic sequence of a Lyme disease spirochaete, Borrelia
burgdorferi.
#cross-references MJD:98065943
#accession E70132
#status preliminary; nucleic acid sequence not shown;
translation not shown
##molecule_type DNA
##residues 1-460 ##label KLF
##cross-references GB:AE000783; NID:g2688160; PID:g2688209;
TIGR:BB0261
#experimental_source strain B31
#superfamily hypothetical: protein BB0261; tetratricopeptide
repeat homology
FEATURE
103-136
#domain tetratricopeptide repeat homology #label TT1\
137-170
#domain tetratricopeptide repeat homology #label TT2\
205-238
#domain tetratricopeptide repeat homology #label TT3\
239-272
#domain tetratricopeptide repeat homology #label TT4\
278-310

```

```

311-344
SUMMARY
#atypical #label TT5\
#domain tetratricopeptide repeat homology #label TT6
#length 460 #molecular-weight 54024 #checksum 9100
Query Match 66.7%; Score 48; DB 1; Length 460;
Best Loca: Similarity 77.8%; Pred. No. 1.30e+01;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
DB 23 IALKLTIGK 31
1111111111
QY 1 ISKLTSCK 9
RESULT 13
ENTRY
#type complete
#major structural core protein - rice dwarf virus
#formal_name rice dwarf virus
#host Oryza sativa (rice)
#16-Jul-1999
#30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change
A45341: S12826; S12821
REFERENCE
A45341
#authors Suzuki, N.; Watanabe, Y.; Kusano, T.; Kitagawa, Y.
#journal Virology (1990) 179:455-459
#title Sequence analysis of the rice dwarf phyto-reovirus segment S3
transcript encoding for the major structural core protein
of 114 kDa.
#cross-references MJD:91021050
#accession A45341
##molecule_type mRNA
##residues 1-1019 ##label SUZ
##cross-references GB:X54520; NID:g61457; PID:CAA38440.1; PID:g61449
S12826
REFERENCE
S12826
#authors Yamada, N.; Uyeda, I.; Kudo, H.; Shikata, E.
#journal Nucleic Acids Res. (1990) 18:6419
#title Nucleotide sequence of rice dwarf virus genome segment 3.
#cross-references MJD:91057125
#accession S12826
##molecule_type genomic RNA
##residues 1-303, 'S', 305-1019 ##label YAM
##cross-references EMBL:D00607; NID:g222500; PID:BAAC0482.1.
PID:d1000937; PID:g222501
S12621
REFERENCE
#authors Kano, H.; Koizumi, M.; Noda, H.; Mizuno, H.; Tsuchihara, T.;
Ishikawa, K.; Hibino, H.; Gmura, T.
#journal Nucleic Acids Res. (1990) 18:6700
#title Nucleotide sequence of rice dwarf virus (RDV) genome segment
S3 coding for 114 k major core protein.
#cross-references MJD:91067474
#accession S12621
##molecule_type genomic RNA
##residues 1-18, 'Y', 20-182, 'HG', 185-245, 'Y', 248-422, 'N', 424-658,
'Y', 660-809, 'N', 811-909, 'N', 911-1005, 'M', 1007-1019
##label KAN
##cross-references GB:D00693
GENETICS
#map_position segment 3
CLASSIFICATION #superfamily rice dwarf virus major structural core protein
KEYWORDS
#length 1019 #molecular-weight 114290 #checksum 2402
SUMMARY
Query Match 66.7%; Score 48; DB 1; Length 1019;
Best Loca: Similarity 58.3%; Pred. No. 1.30e+01;
Matches 7; Conservative 4; Mismatches 1; Indels 0; Gaps 0;
DB 267 VSLKLTSGIGT 278
1111111111
QY 1 ISKLTSCKIAS 12
RESULT 14
ENTRY
#type complete
#transcription-repair coupling factor (mfd) RP598 - Rickettsia

```

```

proWzekli
#normal_name Rickettsia prowazekii
#sequence_revision 21-Nov-1998 #text_change
#date 21-Nov-1998
#accessions HT1564
#reference HT1564
#authors Andersson, S.G.E.; Zomorodipour, A.; Andersson, J.O.;
Scheritz-Ponten, T.; Alismark, U.C.M.; Podowski, R.M.;
Naeslund, A.K.; Eriksson, A.S.; Winkler, H.H.; Kurian,
C.G.
#journal Nature (1998) 396:133-140
#title The genome sequence of Rickettsia prowazekii and the origin
of mitochondria.
#cross-references MUID:99039499
#accession HT1564
#status preliminary: nucleic acid sequence not shown;
translation not shown
#molecule_type DNA
#residues 1-1120 #label AND
#cross-references GB:AJ235272; GB:AJ235269; NC:G3861033; PID:el342886;
PID:G3861142
#experimental_source strain Madrid E
GENETICS
#gene wfd: RP598
#classification *superfamily transcription-repair coupling protein: DEAD/H
box helicase homology
#length 1120 #molecular_weight 127674 #checksum 1852
Query Match 66.7% Score 48; DB 2; Length 1120;
Best Local Similarity 70.0%; Pred. NO. 1.30e-01;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
DB 675 IRLELASGK1 584
QY 1 ISLKLISGK1 10

```

```

RESULT 15
ENTRY
#type complete
#name Bacillus subtilis
#normal_name Bacillus subtilis
#sequence_revision 05-Dec-1997 #text_change
#date 05-Dec-1997
#accessions B69948
#reference B69948
#authors Kinst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.;
Allen, G.; Avery, J.; Bortone, M.G.; Bessieres, P.;
Bollati, A.; Borchert, S.; Boriss, R.; Boursier, L.; Brans,
A.; Braun, M.; Brizgeli, S.-C.; Bron, S.; Brouillet, S.;
Bruschi, C.V.; Caldwell, B.; Capriano, V.; Carter, N.M.;
Chen, S.-K.; Codani, J.-J.; Connerton, I.F.; Cummings, N.J.;
Daniel, R.A.; Denizot, F.; Devine, K.M.; Duesterhoeft, A.;
Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.;
Fabret, C.; Ferrari, E.; Foulger, D.; Fritz, C.; Fujita,
M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galleron, N.; Ghim,
S.Y.; Glaser, P.; Goffeau, A.; Gollightly, E.J.; Grandi, G.;
Guiseppi, G.; Guy, B.J.; Haga, K.; Haech, J.; Harwood,
C.R.; Henaut, A.; Hilbert, R.; Holsappel, S.; Hosono, S.;
Hullo, M.F.; Itaya, M.; Jones, L.; Joris, B.; Karamata, D.;
Kasahara, Y.; Klaerr-Blanchard, M.; Klein, C.; Kobayashi,
Y.; Koetter, P.; Koningstein, G.; Kroch, S.; Kumano, M.;
Kurita, K.; Lapidus, A.; Lardinois, S.; Lauber, J.;
Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.;
Maueel, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno,
M.; Moestl, D.; Nakai, S.; Noback, M.; Noone, D.; O'Reilly,
M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro,
V.; Pohl, T.M.; Portetelle, D.; Porwolik, S.; Prescott,
A.M.; Presecan, E.; Pujic, P.; Purnelle, B.; Rapoport, G.;
Rey, M.; Reynolds, S.; Rieser, M.; Rivolta, C.; Rocha, E.;
Roche, B.; Rose, M.; Sadale, Y.; Sato, I.; Scanlon, E.;
Schleich, S.; Schroeter, R.; Sciffone, F.; Sekiguchi, J.;
Sekowska, A.; Seror, S.; Serror, P.; Shin, B.S.; Soldo,
B.; Sorckin, A.; Tacconi, E.; Takagi, T.; Takahashi, H.;

```

\*\*\*\*\*

WORLD  
WIDE  
WEB  
WWW

(TM)

\*\*\*\*\*

Release 2.1A John F. Collins, BioComputing Research Unit,  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

Msrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Mon Jun 19 16:03:48 2000: MasPar time 2.83 Seconds  
129.156 Million cell updates/sec  
Total output not generated.

Title: US-09-142-524A-6  
Description: (1-12) from US09142524A.peg  
Perfect Score: 72  
Sequence: 1 ISKLISGXIAS 12

Scoring table: FAM 150  
Gap 15

Searched: 8,657 seqs. 30454573 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: swiss-prot38  
iswissprot

Statistics: Mean 24.176; Variance 23.745; scale 1.018

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description                       | Pred. No. |
|------------|-------|-------------|--------|----|-----------------------------------|-----------|
| 1          | 72    | 100.0       | 514    | 1  | MPA2_CRYJA POSSIBLE POLYGALACTURO | 1.09e-06  |
| 2          | 53    | 73.6        | 316    | 1  | Y248_GAEEL HYPOTHETICAL 33.7 KD P | 2.36e-01  |
| 3          | 51    | 70.8        | 255    | 1  | FECE_ECOLI IRON(III) DICITRATE TR | 7.49e-01  |
| 4          | 49    | 68.1        | 573    | 1  | AMPL_SOLTU CHLOROPLAST AMINOPEPTI | 2.30e-00  |
| 5          | 48    | 66.7        | 376    | 1  | YGM_YEAST HYPOTHETICAL 44.3 KD P  | 3.98e-00  |
| 6          | 48    | 66.7        | 1019   | 1  | VPJ_RDV MAJOR 114 KD STRUCTURA    | 3.98e-00  |
| 7          | 48    | 66.7        | 1120   | 1  | MFD_RICPR TRANSCRIPTION-REPAIR C  | 3.98e-00  |
| 8          | 48    | 66.7        | 1585   | 1  | YQBO_BACSU HYPOTHETICAL 171.0 KD  | 3.98e-00  |
| 9          | 47    | 65.3        | 530    | 1  | UL21_HSVB GENE 40 PROTEIN         | 6.82e-00  |
| 10         | 47    | 65.3        | 572    | 1  | LAC3_THACU LACCASE 3 PRECURSOR (E | 6.82e-00  |
| 11         | 47    | 65.3        | 548    | 1  | PEX6_SCHPO PROBABLE PEROXIN-6     | 6.82e-00  |
| 12         | 46    | 63.9        | 121    | 1  | YF39_PYRHO HYPOTHETICAL PROTEIN P | 1.16e-01  |
| 13         | 46    | 63.9        | 124    | 1  | Y214_METJA PUTATIVE HYDROGENASE E | 1.16e-01  |
| 14         | 46    | 63.9        | 145    | 1  | AGGB_ECOLI AGGB PROTEIN PRECURSOR | 1.16e-01  |
| 15         | 46    | 63.9        | 208    | 1  | FTSQ_STRGR CELL DIVISION PROTEIN  | 1.16e-01  |
| 16         | 46    | 63.9        | 251    | 1  | ZNUC_ECOLI HIGH-AFFINITY ZINC UPT | 1.16e-01  |
| 17         | 46    | 63.9        | 386    | 1  | BLAT_ECOLI BETA-LACTAMASE PRECURS | 1.16e-01  |
| 18         | 46    | 63.9        | 316    | 1  | TALB_ECOLI TRANSALDOLASE B (EC 2  | 1.16e-01  |
| 19         | 46    | 63.9        | 359    | 1  | PST_CRIGR ALPHA-2.8-POLYSIALYLTR  | 1.16e-01  |
| 20         | 46    | 63.9        | 359    | 1  | PST_HUMAN ALPHA-2.8-POLYSIALYLTR  | 1.16e-01  |
| 21         | 46    | 63.9        | 359    | 1  | PST_MOUSE ALPHA-2.8-POLYSIALYLTR  | 1.16e-01  |
| 22         | 46    | 63.9        | 551    | 1  | REFR_VARV RIFAMPICIN RESISTANCE   | 1.16e-01  |
| 23         | 46    | 63.9        | 551    | 1  | REFR_VACCV RIFAMPICIN RESISTANCE  | 1.16e-01  |

|    |    |      |      |   |                                   |          |
|----|----|------|------|---|-----------------------------------|----------|
| 24 | 46 | 63.9 | 537  | 1 | NUSM_STRPU                        | 1.16e-01 |
| 25 | 46 | 63.9 | 836  | 1 | YU4K_CASBL HYPOTHETICAL 94.6 KD P | 1.16e-01 |
| 26 | 46 | 63.9 | 1010 | 1 | Y661_HASIN PROBABLE TONB-DEPENDEN | 1.16e-01 |
| 27 | 46 | 63.9 | 1202 | 1 | DPN_LASCIM PROBABLE DNA POLYMERAS | 1.16e-01 |
| 28 | 46 | 63.9 | 1361 | 1 | YH1_YEAST HYPOTHETICAL 153.8 KD   | 1.16e-01 |
| 29 | 45 | 62.5 | 97   | 1 | FLM_SALIY NEGATIVE REGULATOR OF   | 1.16e-01 |
| 30 | 45 | 62.5 | 97   | 1 | FLM_ECOLI NEGATIVE REGULATOR OF   | 1.16e-01 |
| 31 | 45 | 62.5 | 330  | 1 | CAL_HAI CATHEPSIN S PRECURSOR     | 1.16e-01 |
| 32 | 45 | 62.5 | 352  | 1 | Y456_METJA HYPOTHETICAL PROTEIN M | 1.16e-01 |
| 33 | 45 | 62.5 | 376  | 1 | KTHL_HSV23 THYMIDINE KINASE (EC 2 | 1.16e-01 |
| 34 | 45 | 62.5 | 428  | 1 | RF4_KLUCA RF4 PROTEIN             | 1.16e-01 |
| 35 | 45 | 62.5 | 463  | 1 | TIS_SALPO TYPE I RESTRICTION ENZ  | 1.16e-01 |
| 36 | 45 | 62.5 | 556  | 1 | PMGI_MICCO 2,3-BISPHOSPHOGYCEPAT  | 1.16e-01 |
| 37 | 45 | 62.5 | 559  | 1 | PMGI_TORAC 2,3-BISPHOSPHOGYCEPAT  | 1.16e-01 |
| 38 | 45 | 62.5 | 638  | 1 | DNK_RHILE DNK PROTEIN (HEAT SHO   | 1.16e-01 |
| 39 | 45 | 62.5 | 762  | 1 | SLAP_ACEM CELL SURFACE PROTEIN P  | 1.16e-01 |
| 40 | 45 | 62.5 | 1335 | 1 | XDH_DROME XANTHINE DEHYDROGENASE  | 1.16e-01 |
| 41 | 44 | 61.1 | 269  | 1 | YRHF_ECOLI HYPOTHETICAL ABC TRANS | 3.23e-01 |
| 42 | 44 | 61.1 | 330  | 1 | CL_HAI CALPONIN, ACIDIC ISOFO     | 3.23e-01 |
| 43 | 44 | 61.1 | 384  | 1 | AG_ECOLI AGAS PROTEIN             | 3.23e-01 |
| 44 | 44 | 61.1 | 559  | 1 | YV1_BACSU HYPOTHETICAL 74.3 KD P  | 3.23e-01 |
| 45 | 44 | 61.1 | 1850 | 1 | V1_2_CHICK VITELLOGENIN II PRECUR | 3.23e-01 |

ALIGNMENTS

| RESULT ID | MPA2_CRYJA   | STANDARD:                         | RT: | 514 AA: |
|-----------|--|-----------------------------------|-----|---------|
| AC        | P43212   |                                   |     |         |
| DT        | 01-NOV-1995  | (Rel. 32, Created)                |     |         |
| DT        | 01-NOV-1995  | (Rel. 32, Last sequence update)   |     |         |
| DE        | 01-NOV-1997  | (Rel. 35, Last annotation update) |     |         |
| DE        | POSSIBLE POLYGALACTURONASE PRECURSOR (EC 3.2.1.15) (PG) (PEPTINASE)  |                                   |     |         |
| DE        | (MAJOR POLLEN ALLERGEN CRY J 2) (CRY J II)                           |                                   |     |         |
| OS        | Cryptomeria japonica (Japanese cedar)                                |                                   |     |         |
| OC        | Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;   |                                   |     |         |
| OC        | euphyllopytes; Spermatophyta; Coniferopsida; Coniferales;            |                                   |     |         |
| OC        | Taxodiaceae; Cryptomeria   |                                   |     |         |
| RN        | [1]  |                                   |     |         |
| RP        | SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.                             |                                   |     |         |
| RC        | TISSUE-POLLEN;   |                                   |     |         |
| RX        | MEDLINE: 95010777  |                                   |     |         |
| RA        | Namba M., Kurose M., Torigoe K., Hino K., Taniguchi Y., Fukuda S.    |                                   |     |         |
| RA        | Usui M., Kurimoto M.;  |                                   |     |         |
| RT        | Molecular cloning of the second major allergen, Cry J II, from       |                                   |     |         |
| RT        | Japanese cedar pollen.   |                                   |     |         |
| RL        | FEBS Lett. 353:124-128(1994).  |                                   |     |         |
| RN        | [2]  |                                   |     |         |
| RP        | SEQUENCE FROM N.A.   |                                   |     |         |
| RC        | TISSUE-POLLEN;   |                                   |     |         |
| RX        | MEDLINE: 94271186  |                                   |     |         |
| RA        | Komiyama N., Sone T., Shimizu K., Morikubo K., Kino K.;              |                                   |     |         |
| RT        | cDNA cloning and expression of Cry J II the second major allergen of |                                   |     |         |
| RT        | Japanese cedar pollen.   |                                   |     |         |
| RL        | Biochem. Biophys. Res. Commun. 201:1021-1028(1994).                  |                                   |     |         |
| RN        | [3]  |                                   |     |         |
| RP        | SEQUENCE OF 55-68.   |                                   |     |         |
| RX        | MEDLINE: 90342988  |                                   |     |         |
| RA        | Sakaguchi M., Inouye S., Tanai M., Ando S., Usui M., Matukasi I.;    |                                   |     |         |
| RT        | Identification of the second major allergen of Japanese cedar        |                                   |     |         |
| RT        | pollen.  |                                   |     |         |
| RL        | Allergy 45:309-312(1990).  |                                   |     |         |
| CC        | !- CATALYTIC ACTIVITY: RANDOM HYDROLYSIS OF 1,4-ALPHA-D-             |                                   |     |         |
| CC        | GALACTOSIDURONIC LINKAGES IN PECTATE AND OTHER GALACTURONANS.        |                                   |     |         |
| CC        | !- SUBCELLULAR LOCATION: SECRETED OR AMYLOPLAST (POTENTIAL).         |                                   |     |         |
| CC        | !- SIMILARITY: BELONGS TO FAMILY 28 OF GLYCOSYL HYDROLASES           |                                   |     |         |
| CC        | (POLYGALACTURONASES).  |                                   |     |         |

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial

YOUNG  
TENN

10Y . . . . .

laboration-  
station its  
s ee no way  
commercial  
audience/

sé 3:







CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
 CC EMBL: C32216; BAA06447.1; .  
 CC DR EMBL: D84432; BAA12411.1; .  
 CC DF EMBL: 299117; CAB14544.1; .  
 CC DE SUBMITTER: BG11285; YQBO.  
 CC DR PFAM: PF01464; SUT.1.  
 CC KW Hypothetical protein.  
 CC SQ SEQUENCE 1585 AA: 171030 MW: 1F82AB7FC50735F CRC64:

Query Match 66.7% Score 49: DB 1: Length 1585:

Best Local Similarity 70.0% Pred. NO. 3.98e-00:

Matches 7: Conservative 2: Mismatches 1: Indels 0: Gaps 0:

DL 325 LKLTSGKIAN 135

QY 3 LKLTSGKIAS 12

# RESULT 9

ID UL21 HSVB STANDARD: PRT: 530 AA.

AC P28672:

DI 01-DEC-1992 (Rel. 24, Created)

DI 01-DEC-1992 (Rel. 24, Last sequence update)

DI 01-DEC-1992 (Rel. 24, Last annotation update)

DE GENE 40 PROTEIN.

UN 40.

CC Equine herpesvirus type 1 (strain Abap) (EHV-1).

CC Viruses: dsDNA viruses, no RNA stage: Herpesviridae;

CC Alphaherpesvirinae: Varicellovirus.

CC (1)

CC SEQUENCE FROM N.A.

CC MEDLINE: 92295566.

CC Telford E.A.R., Watson M.S., McBride K., Davison A.J.:

CC "The DNA sequence of equine herpesvirus-1."

CC Virology 189:304-316(1992)

CC -- SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV-1 JUL21,

CC EHV-1 4C, EHV-4 JUL21, AND VZV 38.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
 CC EMBL: M86654; AAB02476.1; .  
 CC DR PIR: F36799; WZBD4.  
 CC SQ SEQUENCE 530 AA: 57915 MW: 2E43565AB297CD18 CRC64:

Query Match 65.3% Score 47: DB 1: Length 530:

Best Local Similarity 65.7% Pred. NO. 6.82e-00:

Matches 8: Conservative 2: Mismatches 2: Indels 0: Gaps 0:

DB 126 VSKVTSGIIVS 137

QY 1 LKLTSGKIAS 12

# RESULT 10

ID LAC3\_THACU STANDARD: PRT: 572 AA.

AC Q02079:

DI 01-NOV-1997 (Rel. 35, Created)

DI 01-NOV-1997 (Rel. 35, Last sequence update)

DI 01-NOV-1997 (Rel. 35, Last annotation update)

DE LACCASE 3 PRECURSOR (EC 1.10.3.2) (BENZENEDIOL-OXYGEN OXIDOREDUCTASE)

DE (URUSHIOL OXIDASE) (CIPHENOL OXIDASE).

CC LCC3.

CC Thanatephorus cucumeris (Black scurf of potato) (Rhizoctonia solani).

CC Eukaryota: Fungi: Basidiomycota: Hymenomycetes: Ceratobasidiales;

Ceratobasidiaceae: anamorphic Ceratobasidiaceae: Rhizoctonia

(1)

RN SEQUENCE FROM N.A.

RP STRAIN-RS22:

RX MEDLINE: 96171523.

RA Wahleithner J.A., Xu F., Brown K.M., Brown S.H., Golightly E.J.,

RA Halkier T., Kauppinen S., Pedersen A., Schneider P.:

RA "The identification and characterization of four laccases from the

RA plant pathogenic fungus *Rhizoctonia solani*."

RL Curr. Genet. 29:395-403(1996).

CC -- FUNCTION: LIGNIN DEGRADATION AND DETOXIFICATION OF LIGNIN-DERIVED

CC PRODUCTS (PROBABLE).

CC -- CATALYTIC ACTIVITY: 4 BENZENEDIOL - O(2) - 4 BENZOSEMIQUINONE +

CC 2 H(2)O.

CC -- COFACTOR: BINDS 4 CU-IONS PER MOLECULE. THREE DISTINCT CU

CC CENTERS KNOWN AS TYPE 1 OR BLUE, TYPE 2 OR NORMAL, AND TYPE

CC 3 OR COUPLED BINUCLEAR (BY SIMILARITY).

CC -- SUBUNIT: HOMODIMER (BY SIMILARITY).

CC -- SUBCELLULAR LOCATION: SECRETED.

CC -- TISSUE SPECIFICITY: IN MYCELIA, AT A LOWER LEVEL THAN LCC4.

CC -- SIMILARITY: BELONGS TO THE FAMILY OF MULTICOPPER OXIDASES.

CC -- SIMILARITY: CONTAINS 3 PLASTOCYANIN-LIKE DOMAINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
 CC EMBL: Z54215; CAA90942.1; .

CC PROSITE: PS00079; MULTICOPPER\_OXIDASE1; 1.

CC PROSITE: PS00080; MULTICOPPER\_OXIDASE2; FALSE\_NEG.

CC PFAM: PF00394; Cu-oxidase; 2.

CC Oxidoreductase: Signal: Copper Metal-binding: Glycoprotein: Repeat:

CC Lignin degradation: Multidomain: Polymorphism.

CC SIGNAL 18

CC CHAIN 19 572

CC DOMAIN 21 145

CC DOMAIN 157 304

CC DOMAIN 422 540

CC METAL 82 82

CC METAL 84 84

CC METAL 127 127

CC METAL 129 129

CC METAL 470 470

CC METAL 473 473

CC METAL 475 475

CC METAL 522 522

CC METAL 523 523

CC METAL 524 524

CC METAL 528 528

CC METAL 7

CC CARBOHYD 182

CC CARBOHYD 228

CC CARBOHYD 294

CC CARBOHYD 367

CC CARBOHYD 405

CC VARIANT 159

CC VARIANT 359

CC VARIANT 418

CC VARIANT 448

CC SEQUENCE 572 AA: 63747 MW: DF1E2AF9FF108CE4 CRC64:

Query Match 65.3% Score 47: DB 1: Length 572:

Best Local Similarity 54.5% Pred. NO. 6.82e-00:

Matches 6: Conservative 5: Mismatches 0: Indels 0: Gaps 0:

DB 25 NLKISNGKIAP 35

QY 2 SKLTSGKIAS 12

```

RESULT 11
ID PEX6_SCHPO STANDARD: PRT: 946 AA.
AC G13764
DT 15-JUL-1998 (Rel. 36, Created)
DI 15-JUL-1998 (Rel. 36, Last sequence update)
DR 15-JUL-1998 (Rel. 36, Last annotation update)
DE PROBABLE PEROXIN-6.
GN SPAC1A5.01.
CS Schizosaccharomyces pombe (Fission yeast).
CC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetales;
CC Schizosaccharomycetaceae; Schizosaccharomycetes.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972.
RA Devlin K., Churcher C.M., Barrell B.G., Rajandream M.A., Wood V.:
RE Submitted (Aug-1997) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: ESSENTIAL FOR THE BIOGENESIS OF PEROXISOMES (BY
CC -!- SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE AAA FAMILY OF ATPASES.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: Z98849; CAB1501.1.
CC PROSITE: PSC674; AAA; FALSE NEG.
CC PFAM: PF00041; AAA; 2.
CC PEROXISOME: ATP-binding.
CC NP_BIND: 695 702 ATP (POTENTIAL).
CC SEQUENCE 946 AA: 106505 MW: 140505 MW: F41DF7DE6D31B5 CRC64:
Query Match 65.3% Score 47: DB 1: Length 948:
Best Local Similarity 66.7% Pred. No. 1:16e+01:
Matches 6: Conservative 2: Mismatches 1: Indels 0: Gaps 0:
Db 220 ISRLS03K 228
| | | | |
OY 1 ISLTSJK 9

RESULT 12
ID YF39_PVRHG STANDARD: PRT: 121 AA.
AC Q74C17
DI 15-FEB-2000 (Rel. 39, Created)
DI 15-FEB-2000 (Rel. 39, Last sequence update)
DI 15-FEB-2000 (Rel. 39, Last annotation update)
DE HYPOTHETICAL PROTEIN PH1539.
GN PH1539.
CS Pyrococcus horikoshii.
CC Archaea; Euryarchaeota; Thermococcales; Pyrococcus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O73.
RX MEDLINE: 98344137.
RA Kawarabayashi Y., Sawada M., Horikawa H., Haikawa Y., Hiro Y.,
RA Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,
RA Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y.,
RA Funahashi T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A.,
RA Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,
RA Masuchi Y., Shizuya H., Kikuchi H.:
RI "Complete sequence and gene organization of the genome of a hyper-
RI thermophilic archaeobacterium, Pyrococcus horikoshii O73.",
RI DNA Res. 5:55-76(1998).
CC -!- SIMILARITY: BELONGS TO THE UPF0099 FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

```

```

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AP000006; BAA30649.1.
CC KW Hypothetical protein.
SO SEQUENCE 121 AA: 13377 MW: 167124443A5F98B CRC64:
Query Match 63.9% Score 46: DB 1: Length 121:
Best Local Similarity 60.0% Pred. No. 1:16e+01:
Matches 6: Conservative 2: Mismatches 1: Indels 0: Gaps 0:
Db 17 LKLSKGKIAA 26
| | | | |
OY 3 LKLISGKIAS 12

RESULT 13
ID Y214_METZA STANDARD: PRT: 124 AA.
AC Q57657
DI 01-NOV-1997 (Rel. 35, Created)
DI 01-NOV-1997 (Rel. 35, Last sequence update)
DI 01-NOV-1997 (Rel. 35, Last annotation update)
DE PUTATIVE HYDROGENASE EXPRESSION/FORMATION PROTEIN MJ0214.
GN MJ0214.
OS Methanococcus jannaschii.
CC Archaea; Euryarchaeota; Methanococcales; Methanococcus.
CC Methanococcus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067.
RX MEDLINE: 96337999.
RA But C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald C.M., Clayton R.A., Soukey J.D.,
RA Kerlavage A.R., Dougherty B.A., Iont J.-P., Adams M.D., Reich C.L.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick C.M., Glöck A.,
RA Scott J.C., Geohagen N.S.M., Weidman J.F., Fainman J.L., Nguyen D.,
RA Uitterback T.R., Kelley J.M., Emerson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurt M.A., Kaine B.P., Horodovskiy M.,
RA Klek H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.:
RI "Complete genome sequence of the methanogenic archaeon, Methanococcus
RI jannaschii."
RI Science 273:1058-1073(1996).
CC -!- SIMILARITY: BELONGS TO THE HYPA/HYPE FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: J67477; AAB98198.1.
CC TIGR: MJ0214.
DR PROSITE: PS01249; HYPA: 1.
DR PFAM: PF01155; HYPA: 1.
KW Hypothetical protein; Zinc-finger.
FT 2N-FING 78 100 C4-TYPE (POTENTIAL).
SO SEQUENCE 124 AA: 14058 MW: 176980505A96957AD CRC64:
Query Match 63.9% Score 46: DB 1: Length 124:
Best Local Similarity 66.7% Pred. No. 1:16e+01:
Matches 6: Conservative 2: Mismatches 1: Indels 0: Gaps 0:
Db 103 IRKLSSGK 111
| | | | |
OY 1 ISLKLTSJK 9

RESULT 14
ID AGGB_ECOLI STANDARD: PRT: 145 AA.

```

```

AC P46006;
DI 01-NOV-1995 (Rel. 32, Created)
DI 01-NOV-1995 (Rel. 32, Last sequence update)
DI 01-NOV-1997 (Rel. 35, Last annotation update)
DE AGGB PROTEIN PRECURSOR.
UN AGGB.
OS Escherichia coli.
OC Plasmid P17-2.
OC Bacteria: Proteobacteria: gamma subdivision: Enterobacteriaceae:
OC Escherichia.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ENTEROAGGREGATIVE 17-2 / SEROTYPE O3:H2;
RX MEDLINE: 94327462.
RA Savarino S.J., Fox P., Deng Y., Nataro J.P.;
RT Identification and characterization of a gene cluster mediating
RT enteroaggregative Escherichia coli aggregative adherence fimbria I
R1 biogenesis.
RL J. Bacteriol. 176:4949-4957(1994).
CC -!- FUNCTION: NOT KNOWN.
CC -!- SIMILARITY: HIGH, TO E. COLI AFAD PROTEIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: C12894; AAA57453.1;
CC Fimbrin; Plasmid; Signal.
CC SIGNAL 1 24
CC CHAIN 25 145 AGGB PROTEIN.
CC SEQUENCE 145 AA: 15837 MW: 40380CF0526F3DBF CRC64;

Query Match 63.9%; Score 46; DB 1; Length 145;
Best Local Similarity 50.0%; Pred. No. 1:136-01;
Matches 5; Conservative 5; Mismatch 0; Indels 0; Gaps 0;

DB 42 GKXKATGRIA 51
QY 2 SKLTSCKIA 11

RESULT 15
ID FTSG_STRGR STANDARD; PRI: 208 AA.
AC P45503;
DI 01-NOV-1995 (Rel. 32, Created)
DI 01-NOV-1995 (Rel. 32, Last sequence update)
DE CELL DIVISION PROTEIN FTSG HOMOLOG (FRAGMENT).
GN FTSG.
OS Streptomyces griseus.
OC Bacteria: Firmicutes: Actinobacteria: Actinobacteridae.
OC Actinomycetales: Streptomycetaceae: Streptomyces.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NRRL B2682;
RX MEDLINE: 94374704.
RA Dharmatilake A.J., Kendrick K.E.;
RT Expression of the division-controlling gene ftsZ during growth and
RT sporulation of the filamentous bacterium Streptomyces griseus.
RL Gene 147:21-28(1994).
CC -!- FUNCTION: THIS PROTEIN MAY BE INVOLVED IN ANOMALOUS FILAMENT
CC GROWTH. MAY INTERACT WITH FTSG, AND PBP3 (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN (BY SIMILARITY).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```

---

```

CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U07344; AAA56888.1;
CC Cell division.
KW NON_TER 1
FT DOMAIN <1 208
SQ SEQUENCE 208 AA: 21942 MW: 7009AB6699746E82 CRC64;

Query Match 63.9%; Score 45; DB 1; Length 208;
Best Local Similarity 41.7%; Pred. No. 1:136-01;
Matches 5; Conservative 4; Mismatch 3; Indels 0; Gaps 0;

DB 157 ISLRLTRDRVVT 168
QY 1 ISLKLTSCKIAS 12

Search completed: Mon Jun 19 16:03:54 2000
Job time : 6 secs.

```



Matches 7: Conservative 4: Mismatches 1: Indels 0: Gaps 0:

DE 209 MSKMGTKIAA 220  
 QY 1 ISLKLTSKIAS 12

RESULT 2  
 ID Q33358 PRELIMINARY: PRT: 182 AA.  
 AC Q33358  
 DT 01-JAN-1998 (TrEMBLrel. 05, Created)  
 DI 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)  
 DE 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)  
 DE HYPOTHETICAL 19.4 KD PROTEIN.  
 GN MTC260.0.033.  
 OS Mycobacterium tuberculosis.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-H37RV.  
 RA BROWN D., CHURCHER C.M.  
 RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.  
 RN (2)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-H37RV.  
 RA PARKHILL J., BARRELL B.G., RAJANDREAM M.A.  
 RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.  
 RN (3)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-H37RV.  
 RA MEDLINE; 96161548.  
 RA PHILIPP W.J., POULET S., FIEGLMEIER K., PASCOPELLA L.,  
 RA BALASUBRAMANIAN V., HEYM B., BERGH S., BLOOM B.R., JACOBS W.R. JR.,  
 RA COLE S.T.  
 RL "An integrated map of the genome of the tubercle bacillus.  
 RI Mycobacterium tuberculosis H37RV, and comparison with Mycobacterium  
 RI leprae."  
 RC Proc. Natl. Acad. Sci. U.S.A. 94:3132-3137(1996).  
 DR EMBL: Z97811; GAB0750.1;  
 KW Hypothetical protein.  
 SQ SEQUENCE 182 AA: 19401 MW: 25881D7F CRC32:

Query Match 72.2% Score 52: DB 2: Length 182:  
 Best Local Similarity 63.5% Pred. No. 1.05e+00:  
 Matches 7: Conservative 4: Mismatches 1: Indels 0: Gaps 0:

DE 131 ISLKLTSKIA 14  
 QY 1 ISLKLTSKIA 11

RESULT 3  
 ID Q92W11 PRELIMINARY: PRT: 623 AA.  
 AC Q92W11

DT 01-MAY-1999 (TrEMBLrel. 10, Created)  
 DI 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)  
 DE 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)  
 DE PUTATIVE RECEPTOR-LIKE PROTEIN KINASE.  
 GN F16P2.37.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;  
 OC core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;  
 OC Arabidopsids.  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CV. COLUMBIA.  
 RA ROUNSLEY S.D., LIN X., KAUL S., SHEA T.P., FUJII C.Y., MASON T.M.,  
 RA SHEN M., RONNING C.M., FRASER C.M., SOMERVILLE C.R., VENTER J.C.;  
 RI "Arabidopsis thaliana chromosome II BAC F16P2 genomic sequence."  
 RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AC004561; AAC95210.1;  
 KW

DR HSSP: P02867; IRIN.  
 KW Receptor; Kinase.  
 SQ SEQUENCE 623 AA: 68216 MW: F0DAD5D7 CRC32:

Query Match 70.8% Score 51: DB 10: Length 623:  
 Best Local Similarity 77.8% Pred. No. 1.83e+00:  
 Matches 7: Conservative 2: Mismatches 0: Indels 0: Gaps 0:

DE 185 ISLRLASOK 193  
 QY 1 ISLKLISOK 9

RESULT 4  
 ID Q92W09 PRELIMINARY: PRT: 627 AA.  
 AC Q92W09  
 DT 01-MAY-1999 (TrEMBLrel. 10, Created)  
 DI 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)  
 DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)  
 DE PUTATIVE RECEPTOR-LIKE PROTEIN KINASE.  
 GN F16P2.40.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;  
 OC core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;  
 OC Arabidopsids.  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CV. COLUMBIA.  
 RA ROUNSLEY S.D., LIN X., KAUL S., SHEA T.P., FUJII C.Y., MASON T.M.,  
 RA SHEN M., RONNING C.M., FRASER C.M., SOMERVILLE C.R., VENTER J.C.;  
 RI "Arabidopsis thaliana chromosome II BAC F16P2 genomic sequence."  
 RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AC004561; AAC95213.1;  
 DR HSSP: P03045; 13JQ.  
 KW Receptor; Kinase.  
 SQ SEQUENCE 627 AA: 68788 MW: 7D6CCA4D CRC32:

Query Match 70.8% Score 51: DB 10: Length 627:  
 Best Local Similarity 77.8% Pred. No. 1.83e+00:  
 Matches 7: Conservative 2: Mismatches 0: Indels 0: Gaps 0:

DE 185 ISLRLASOK 193  
 QY 1 ISLKLISOK 9

RESULT 5  
 ID Q92HX5 PRELIMINARY: PRT: 157 AA.  
 AC Q92HX5  
 DT 01-MAY-1999 (TrEMBLrel. 10, Created)  
 DI 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)  
 DT 01-MAY-1999 (TrEMBLrel. 10, Last annotation update)  
 DE P2 ISBM1.  
 OS Brucella melitensis.  
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
 OC Brucella.  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-16M;  
 RA GODFROID F., TAMINIAU B., DANESI I., DENOEL P., TIBOR A., WEYNANTS V.,  
 RA CLOECKAERT A., GODFROID J., LETESSON J.J.;  
 RI "Identification of the perosamine synthetase gene of Brucella  
 RI melitensis 16M and involvement of lipopolysaccharide O side chain in  
 RI Brucella survival in mice and in macrophages."  
 RL Infect. Immun. 66:0-0(1998).  
 DR EMBL: AF047478; AAC98623.1;  
 SQ SEQUENCE 157 AA: 17565 MW: 52393578 CRC32:

Query Match 69.4% Score 50: DB 2: Length 157:  
 Best Local Similarity 54.5% Pred. No. 3.17e+00:  
 Matches 5: Conservative 4: Mismatches 1: Indels 0: Gaps 0:

```

43 VNRLTGGQIA 53
1 ISKLTSCKIA 12

RESULT 6
AC 028993 PRELIMINARY: PRI: 407 AA.
DI 01-JAN-1998 (TREMblrel. 05, Created)
DI 01-JAN-1998 (TREMblrel. 05, Last sequence update)
DI 01-NOV-1998 (TREMblrel. 08, Last sequence update)
DI 01-NOV-1998 (TREMblrel. 08, Last annotation update)
DE CONSERVED HYPOTHETICAL PROTEIN.
GN AF1278.
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobales; Archaeoglobaceae;
OC Archaeoglobus.
RN [1]
RC SEQUENCE FROM N.A.
RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX MEDLINE: 98049343.
RA KLENK H.-P., CLAYTON R.A., TOMB J.-F., WHITE O., NELSON K.E.,
RA KETCHUM K.A., DODSON R.J., CHINN M., HICKEY E.K., PETERSON J.D.,
RA RICHARDSON D.L., KERLAVAGE A.R., GRAHAM D.E., KYRPIDES N.C.,
RA FLEISCHMANN R.D., QUACKENBUSH J., LEE N.H., SUTTON G.G., GILL S.,
RA KIRKNESS E.F., DOUGHERTY B.A., MCKENNEY K., ADAMS M.D., LOFTUS B.,
RA PETERSON S., REICH C.I., MCNEIL L.K., BADGER J.H., GLODEK A., ZHOU L.,
RA OVERBECK R., GPCAYNE J.D., WEIDMAN J.F., MCDONALD L., UTTERBACK T.,
RA COTTON M.D., SPRIGGS T., ARTIACH P., KAINE B.P., SYKES S.M.,
RA SADOW P.W., D'ANDREA K.P., BOWMAN C., FUJII C., GARLAND S.A.,
RA MASON T.M., OLSEN G.J., FRASER C.M., SMITH H.C., WOESE C.R.,
RA VENTER J.C.;
RI "The complete genome sequence of the hyperthermophilic, sulphate-
RI reducing archaeon Archaeoglobus fulgidus."
RL Nature 390:364-370(1997).
RX EMBL: AE301016; AB96967.1;
RX TIGR: AF1278.
KW Hypothetical protein.
SQ SEQUENCE 407 AA: 44931 MW: E38F5477 CRC32:

Query Match 68.1%; Score 49; DB 1; Length 407;
Best Local Similarity 77.6%; Pred. NO. 5.41e+00;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

DB 193 LKMGKQIA 201
1 ISKLTSCKIA 12
3 LKLTSGKIA 11

RESULT 7
AC 032149 PRELIMINARY: PRI: 412 AA.
DI 01-JAN-1998 (TREMblrel. 05, Created)
DI 01-JAN-1998 (TREMblrel. 05, Last sequence update)
DI 01-NOV-1998 (TREMblrel. 08, Last sequence update)
DE YUBH PROTEIN.
GN YUBH.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
RN [1]
RC SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE: 98044033.
RA KUNST F., OGASAWARA N., MOSZER I., ALBERTINI A.M., ALLONI G.,
RA AZEVEDO V., BERTERO M.G., BESSIERES P., BOLOIN A., BORCHERT S.,
RA BORRIS R., BOURSIER L., BRANS A., BRAUN M., BRIGNELL S.C., BRON S.,
RA BROUILLET S., BRUSCHI C.V., CALDWELL B., CAPUANO V., CARTER N.M.,
RA CHOI S.K., CODANI J.J., CONNERTON I.F., CUMMINGS N.D., DANIEL R.A.,
RA DENIZOT F., DEVINE K.M., DUNSTON J., EHRICH S.D., EMMERSON P.T.,
RA ENTIAN K.D., ERRINGTON J., FABRET C., FERRARI E., FOULGER D.,
RA FRITZ C., FUJITA M., FUJITA Y., FUMA S., GALI221 A., GALLERON N.,
RA GHIM S.Y., GLASER P., GORFEAU A., GOLIGHTLY E.J., GRANDI G.,
RA GUISSEPI G., GUY B.C., HAGA K., HAIBCH J., HARWOOD C.R., HENAUT A.,

```

```

RA HILBERT H., HOLSAPPEL S., ROS S.S., HULLO M.F., ITAYA M., JONES L.,
RA JORIS B., KARAMATA D., KASAHARA Y., KLAERR-BLANCHARD M., KLEIN C.,
RA KOBAYASHI Y., KOETTES P., KONIGSTEIN G., KROGH S., KUMANG M.,
RA KURITA K., LAPIDUS A., LARDIN S.S., LAUBER J., LAZAREVIC V.,
RA LEE S.M., LEVINE A., LIU H., MASUDA S., MAUEL C., MEDIGUE C.,
RA MEDINA N., MELLADO R.P., MIZUNO M., MOEST D., NAKAI S., NISACK M.,
RA NOONE D., O'REILLY M., OGAWA K., OGIMURA A., ODEGA B., PARK S.H.,
RA PARRO V., POHL T.M., PORTETELLE D., PORWOLLIK S., PRESCOTT A.M.,
RA PRESECAN E., PUJIC P., PURNELLE B., RAPOPORT G., REY M., REYNOLDS S.,
RA RIEGER M., RIVOLTA C., ROCHA E., ROCHE B., ROSE M., SADAIE Y.,
RA SAITO T., SCANLAN E., SCHLEICH S., SCHROETER R., SCOFFONE F.,
RA SEKIGUCHI J., SEKOWSKA A., SEF R.S., SERROR P., SHIN B.S., SILDON B.,
RA SOROKIN A., TACCONI E., TAKAGI T., TAKAHASHI H., TAKEMARU K.,
RA TAKEUCHI M., TAKAKOSHI A., TAKAKA T., TERPSTRA P., TOGNONI A.,
RA TOSATO V., UCHIYAMA S., VANDER M., VANNIER F., VASSAROTTI A.,
RA VIARI A., WAMBUTT R., WEDLER H., WEITZENESGER T., YATA K.,
RA WINTERS P., WIPUT A., YAMAMOTO T., YAMANE K., YASUMOTO K.,
RA YOSHIDA K., YOSHIKAWA H.F., YOSHIDA E., YOSHIKAWA H., YANCHIN A.,
RI "The complete genome sequence of the gram-positive bacterium Bacillus
RI subtilis."
RL Nature 390:249-256(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA KUNST F., OGASAWARA N., YOSHIKAWA H., DANCHIN A.;
RI Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL: Z99120; CAB15243.1;
SQ SEQUENCE 412 AA: 45519 MW: 7F26E2F0 CRC32:

Query Match 68.1%; Score 49; DB 2; Length 412;
Best Local Similarity 60.0%; Pred. NO. 5.44e+00;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

DB 263 LRLTCKKITA 272
1 ISKLTSCKIA 12
3 LKLTSGKIA 12

RESULT 8
ID 082633 PRELIMINARY: PRI: 210 AA.
AC 082633;
DI 01-NOV-1998 (TREMblrel. 08, Created)
DI 01-NOV-1998 (TREMblrel. 08, Last sequence update)
DI 01-NOV-1998 (TREMblrel. 08, Last annotation update)
DE HYPOTHETICAL 23.5 KD PROTEIN
GN F25P21.20.
OS Arabidopsis thaliana (Mouse; Arabidopsis).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Rosidae; eurosids 1; Brassicales; Brassicaceae;
OC Arabidopsis.
RN [1]
RP SEQUENCE FROM N.A.
RA BEVAN M., TERRY N., ARDILES W., BUYSSHAERT C., DASSEVILLE R.,
RA DE CLERCK R., DE KENSER A., NEY P., ROUZE P., VAN DEN DAELE H.,
RA VILLARDEL R., GIELEN J., VAN MONTAGU M., HOEISEL J., MEWES H.W.,
RA MAYER K.F.X., SCHUELLER C.;
RI Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA EU ARABIDOPSIS SEQUENCING PROJECT;
RI Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL031804; CAA21199.1;
KW Hypothetical protein.
SQ SEQUENCE 210 AA: 23517 MW: 7AF75A82 CRC32:

Query Match 56.7%; Score 48; DB 10; Length 210;
Best Local Similarity 70.0%; Pred. NO. 9.27e+00;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

DB 89 LKMTCKKIAS 98
1 ISKLTSCKIA 12
3 LKLTSGKIA 12

```

Best Local Similarity 77.8%: Prod. No. 9.27e-00;  
Matches 7: Conservative 1: Mismatches 1: Indels 0: Gaps 0;

DB 23 ISKLTSGK 31  
QY 1 ISKLTSGK 9

RESULT 1:  
ID Q98630 PRELIMINARY: PRI: 1019 AA.  
AC Q98630:  
DI 01-FEB-1997 (TrEMBLrel. 02, Created)  
DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)  
DE 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)  
DE MAJOR CORE PROTEIN  
DE RICE DWARF VIRUS (RDV)  
DE Viruses: GRANA VIRUSES; Reoviridae: Fibriovirus.  
DE [1]  
RN SEQUENCE FROM N.A.  
RP STRAIN-CHINESE  
RX MEDLINE: 91021050.  
RA SUZUKI N., WATANABE Y., KUSAN T., KITAGAWA Y.  
RT "Sequence analysis of the rice dwarf phytochrome segment N"  
RT "Transcript encoding for the major structural core protein of the"  
RT Kda.:"  
RL Virg-ogy 179:455-459(1990).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-CHINESE  
RA ZHANG F., LI Y., LIU Y., AN C., CHEN Z.  
RL Submitted (SEP-1996) to the EMBL/GenBank/DDBJ databases.  
DR EMBL: U72757; AAB17618.1;  
SQ SEQUENCE 1019 AA: 114102 MW: 1100587 CRC12:

Query Match 66.7%: Score 48: DB 14: Length 1019;  
Best Local Similarity 58.3%: Prod. No. 9.27e-00;  
Matches 7: Conservative 4: Mismatches 1: Indels 0: Gaps 0;

DB 267 VSLKLTSGK 278  
QY 1 ISKLTSGK 12

RESULT 12:  
ID Q19004 PRELIMINARY: PRI: 1743 AA.  
AC Q19004:  
DI 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DE 01-MAY-1999 (TrEMBLrel. 01, Last annotation update)  
DE SMILAR TO GUCAN  
DE D2095.3  
CN Caenorhabditis elegans.  
CC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabdita: Rhabditidae;  
CC Rhabdita: Rhabditidae; Rhabditidae; Pelodidae; Caenorhabditidae.  
RN [1]  
RX MEDLINE: 94150718.  
RA WILSON R., AINSWORTH R., ANDERSON K., BAYNES C., BERKS M.,  
RA BONFIELD J., BURTON C., CONNELL M., COPSEY T., COOPER J., COULSON A.,  
RA GRANTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,  
RA GARDNER A., GREEN P., HARKINS T., HILLIER L., JIER M., JOHNSTON L.,  
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,  
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,  
RA PARSONS J., PERCY C., RIEKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,  
RA SHALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,  
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,  
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.,  
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
RT elegans.";  
RL Nature 368:32-38(1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA GEISEL C.;

RESULT 9:  
ID Q51695 PRELIMINARY: PRI: 309 AA.  
AC Q51695:  
DI 01-JUN-1998 (TrEMBLrel. 06, Created)  
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)  
DE 01-MAY-1999 (TrEMBLrel. 08, Last annotation update)  
DE ABC TRANSPORTER, ATP-BINDING PROTEIN.  
GN BB0754.  
CS Borrelia burgdorferi (Lyme disease spirochete).  
CC Bacteria: Spirochaetales: Spirochaetaceae; Borrelia.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-ATCC 35210 / H31;  
RX MEDLINE: 98065943.  
RA FRASER C.M., CASSENS S., HUANG W.M., SUTTON G.G., CLAYTON R.A.,  
RA LATHIGRA R., WHITE C., KETCHUM K.A., DODSON R., HICKEY E.K., GWINN M.,  
RA DOUGHERTY B., TOMB J.-F., FLEISCHMANN R.D., RICHARDSON D.,  
RA PETERSON J., KERLAVAGE A.R., QUACKENBUSH J., SALZBERG S., HANSON M.,  
RA VAN VUGT R., PALMER N., ADAMS M.D., GOCAYNE J.D., WEIDMAN J.,  
RA UTTERBACK T., WATTHEY L., McDONALD L., ARTIACH P., BOWMAN C.,  
RA GARLAND S., FULII C., COTTON M.D., HORST K., ROBERTS K., HATCH B.,  
RA SMITH H.O., VENTER J.C.;  
RT "Genomic sequence of a Lyme disease spirochete, Borrelia  
RT burgdorferi.";  
RL Nature 390:580-586(1997).  
DR EMBL: AE001175; AAC67099.1;  
DP TIGR: BB0754;  
CP PFAM: PF00005; ABC\_tran: 1;  
KW ATP-binding.  
SQ SEQUENCE 309 AA: 45183 MW: 33566281 CRC12:

Query Match 96.7%: Score 48: DB 2: Length 309;  
Best Local Similarity 77.8%: Prod. No. 9.27e-00;  
Matches 7: Conservative 1: Mismatches 1: Indels 0: Gaps 0;

DB 262 ISKLTSGK 270  
QY 1 ISKLTSGK 9

RESULT 10:  
ID Q51276 PRELIMINARY: PRI: 460 AA.  
AC Q51276:  
DI 01-JUN-1998 (TrEMBLrel. 06, Created)  
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)  
DE 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)  
DE HYPOTHETICAL 54.0 KD PROTEIN.  
GN BB0261.  
CS Borrelia burgdorferi (Lyme disease spirochete).  
CC Bacteria: Spirochaetales: Spirochaetaceae; Borrelia.  
RN [1]  
RX MEDLINE: 98065943.  
RA FRASER C.M., CASSENS S., HUANG W.M., SUTTON G.G., CLAYTON R.A.,  
RA LATHIGRA R., WHITE C., KETCHUM K.A., DODSON R., HICKEY E.K., GWINN M.,  
RA DOUGHERTY B., TOMB J.-F., FLEISCHMANN R.D., RICHARDSON D.,  
RA PETERSON J., KERLAVAGE A.R., QUACKENBUSH J., SALZBERG S., HANSON M.,  
RA VAN VUGT R., PALMER N., ADAMS M.D., GOCAYNE J.D., WEIDMAN J.,  
RA UTTERBACK T., WATTHEY L., McDONALD L., ARTIACH P., BOWMAN C.,  
RA GARLAND S., FULII C., COTTON M.D., HORST K., ROBERTS K., HATCH B.,  
RA SMITH H.O., VENTER J.C.;  
RT "Genomic sequence of a Lyme disease spirochete, Borrelia  
RT burgdorferi.";  
RL Nature 390:580-586(1997).  
DR EMBL: AE001175; AAC66691.1;  
DP TIGR: BB0261;  
KW Hypothetical protein.  
SQ SEQUENCE 460 AA: 54024 MW: 54143655 CRC12:

Query Match 66.7%: Score 48: DB 2: Length 460;

Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.  
 ID 13  
 AC 028751  
 RA WATERSTON R.

Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.  
 ID 14  
 AC 028752  
 RA WATERSTON R.

Query Match 56.7% Score 48 DB 5 Length 1743  
 Best Local Similarity 72.7% Pred. No. 1.57e-01  
 Matches 8 Conservative 2 Mismatches 1 Indels 0 Gaps 0

DB 944 SKPTSGKIAS 954  
 ID 13  
 AC 028751

Query Match 56.7% Score 47 DB 5 Length 1743  
 Best Local Similarity 72.7% Pred. No. 1.57e-01  
 Matches 8 Conservative 2 Mismatches 1 Indels 0 Gaps 0

Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.  
 ID 15  
 AC 028753  
 RA WATERSTON R.

Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.  
 ID 16  
 AC 028754  
 RA WATERSTON R.

Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.  
 ID 17  
 AC 028755  
 RA WATERSTON R.

Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.  
 ID 18  
 AC 028756  
 RA WATERSTON R.

Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.  
 ID 19  
 AC 028757  
 RA WATERSTON R.

Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.  
 ID 20  
 AC 028758  
 RA WATERSTON R.

Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.  
 ID 21  
 AC 028759  
 RA WATERSTON R.

Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.  
 ID 22  
 AC 028760  
 RA WATERSTON R.

Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.  
 ID 23  
 AC 028761  
 RA WATERSTON R.

Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.  
 ID 24  
 AC 028762  
 RA WATERSTON R.

Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.  
 ID 25  
 AC 028763  
 RA WATERSTON R.

Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.  
 ID 26  
 AC 028764  
 RA WATERSTON R.

Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.  
 ID 27  
 AC 028765  
 RA WATERSTON R.

Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.  
 ID 28  
 AC 028766  
 RA WATERSTON R.

Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.  
 ID 29  
 AC 028767  
 RA WATERSTON R.

Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.  
 ID 30  
 AC 028768  
 RA WATERSTON R.

Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.  
 ID 31  
 AC 028769  
 RA WATERSTON R.

Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.  
 ID 32  
 AC 028770  
 RA WATERSTON R.

Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.  
 ID 33  
 AC 028771  
 RA WATERSTON R.

Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.  
 ID 34  
 AC 028772  
 RA WATERSTON R.

Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.  
 ID 35  
 AC 028773  
 RA WATERSTON R.

Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.  
 ID 36  
 AC 028774  
 RA WATERSTON R.

Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.  
 ID 37  
 AC 028775  
 RA WATERSTON R.

Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.  
 ID 38  
 AC 028776  
 RA WATERSTON R.

Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.  
 ID 39  
 AC 028777  
 RA WATERSTON R.

Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.  
 ID 40  
 AC 028778  
 RA WATERSTON R.

Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.  
 ID 41  
 AC 028779  
 RA WATERSTON R.

Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.  
 ID 42  
 AC 028780  
 RA WATERSTON R.

Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.  
 ID 43  
 AC 028781  
 RA WATERSTON R.

Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.  
 ID 44  
 AC 028782  
 RA WATERSTON R.

Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.  
 ID 45  
 AC 028783  
 RA WATERSTON R.

Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.  
 ID 46  
 AC 028784  
 RA WATERSTON R.

Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.  
 ID 47  
 AC 028785  
 RA WATERSTON R.

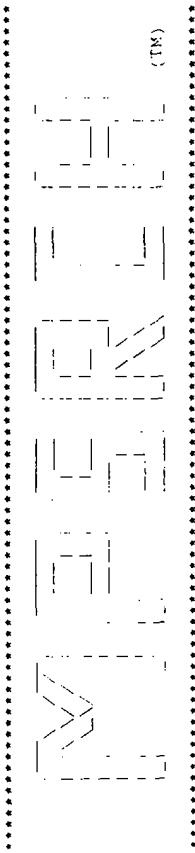
Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.  
 ID 48  
 AC 028786  
 RA WATERSTON R.

Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.  
 ID 49  
 AC 028787  
 RA WATERSTON R.

Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.  
 ID 50  
 AC 028788  
 RA WATERSTON R.



DR EMSL 291565: AA442161: JOINED.  
SQ SEQUENCE 332 AA: 39518 MW: 3DD1SCED CRC32:  
Query Match 65.38: Score 47: DB 5: Length 332:  
Best Local Similarity 45.58: Pred. No. 1.57e+01:  
Matches 5: Conservative 5: Mismatches 1: Indels 0: Gaps 0:  
DB 211: ILKUSGRVP 221  
QY 1: ISLKLISGKIA 11  
Search completed: Mon Jun 19 16:04:24 2000  
Job time : 13 secs.



Release 3.1A John F. Collins, BioComputing Research Unit,  
Copyright (c) 1994-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

KSearch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Mon Jun 19 15:56:22 2000: MasPar time 5.66 Seconds  
439,141 Million cell updates/sec

Facular output not generated.

Files: US-09-142-524A-2  
Description: (1:05) FROM US09142524A.rep  
Perfect Score: 406  
Sequence: 1 MKVTVAFGNPGNRRRVFKR.....FDGRVGGIIAAYONPASWK 105

Scoring table: FAM 150  
Gap 11

Searched: 149963 seqs, 28566186 residues

Post-processing: Minimum Match 08  
Listing first 45 summaries

Database: a-geneseq15  
luceneseqp

Statistics: Mean 29.704; Variance 127.15; scale 0.243

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result | No. | Score | Query | Match | Length | DB     | ID                     | Description | Pred. No. |
|--------|-----|-------|-------|-------|--------|--------|------------------------|-------------|-----------|
| 1      | 1   | 9.6   | 10    | 105   | 105    | W27370 | Multi-epitope peptide  | 4.14e-66    |           |
| 2      | 2   | 9.1   | 6.9   | 134   | 134    | W27372 | Multi-epitope peptide  | 1.86e-42    |           |
| 3      | 3   | 8.7   | 4.8   | 80    | 80     | W27369 | Multi-epitope peptide  | 2.08e-24    |           |
| 4      | 4   | 17.2  | 21.3  | 450   | 450    | R69791 | Japonicum allergen res | 4.23e-06    |           |
| 5      | 5   | 17.2  | 21.3  | 514   | 514    | R53690 | Japanese cedar pollen  | 4.23e-06    |           |
| 6      | 6   | 17.2  | 21.3  | 514   | 514    | R74333 | Japanese cedar pollen  | 4.23e-06    |           |
| 7      | 7   | 17.2  | 21.3  | 514   | 514    | R33509 | Japan cedar pollen cry | 4.23e-06    |           |
| 8      | 8   | 17.2  | 21.3  | 514   | 514    | R69792 | Japonicum allergen     | 4.23e-06    |           |
| 9      | 9   | 17.2  | 21.3  | 514   | 514    | R31586 | Cedar pollen allergen  | 4.23e-06    |           |
| 10     | 10  | 14.9  | 18.5  | 514   | 514    | W43445 | Chamaecyparis obtusa p | 3.96e-04    |           |
| 11     | 11  | 14.9  | 18.5  | 514   | 514    | W42122 | Japanese cypress polle | 3.96e-04    |           |
| 12     | 12  | 14.4  | 17.9  | 353   | 353    | R15187 | Cedar pollen allergen  | 1.05e-03    |           |
| 13     | 13  | 14.4  | 17.9  | 353   | 353    | R75389 | Japanese cedar pollen  | 1.05e-03    |           |
| 14     | 14  | 14.4  | 17.9  | 374   | 374    | R60165 | Japanese cedar pollen  | 1.05e-03    |           |
| 15     | 15  | 14.4  | 17.9  | 374   | 374    | R45541 | Cry j pollen allerg    | 1.05e-03    |           |
| 16     | 16  | 14.4  | 17.9  | 374   | 374    | R62490 | Cry j Japanese Cedar   | 1.05e-03    |           |
| 17     | 17  | 14.4  | 17.9  | 374   | 374    | R19137 | Cry j                  | 1.05e-03    |           |
| 18     | 18  | 13.8  | 17.1  | 17    | 17     | R15181 | Cedar pollen allergen  | 3.32e-03    |           |
| 19     | 19  | 13.7  | 17.0  | 18    | 18     | W60346 | Sugi allergen protein  | 4.02e-03    |           |
| 20     | 20  | 13.7  | 16.3  | 17    | 17     | R15188 | Cedar pollen allergen  | 1.26e-02    |           |
| 21     | 21  | 13.7  | 16.3  | 20    | 20     | W42171 | T-cell epitope peptide | 1.26e-02    |           |
| 22     | 22  | 12.4  | 15.4  | 20    | 20     | R45550 | Cry j pollen allerg    | 4.72e-02    |           |
| 23     | 23  | 12.4  | 15.4  | 20    | 20     | R52499 | Cry j Japanese Cedar   | 4.72e-02    |           |

|    |     |      |     |   |        |                        |          |
|----|-----|------|-----|---|--------|------------------------|----------|
| 24 | 124 | 15.4 | 30  | 1 | 1579   | Cry j j pollen allerg  | 4.72e-02 |
| 25 | 123 | 15.3 | 47  | 1 | 1579   | Sugi allergen protein  | 4.72e-02 |
| 26 | 117 | 14.5 | 14  | 1 | 535    | Japanese cedar pollen  | 1.74e-01 |
| 27 | 117 | 14.5 | 15  | 1 | 497    | Japan cedar pollen mat | 1.74e-01 |
| 28 | 117 | 14.5 | 17  | 1 | 497    | Sugi allergen protein  | 1.74e-01 |
| 29 | 115 | 14.3 | 17  | 1 | R15222 | Cedar pollen allergen  | 2.50e-01 |
| 30 | 114 | 14.1 | 367 | 1 | R45573 | Cry j j                | 3.20e-01 |
| 31 | 114 | 14.1 | 370 | 1 | R45578 | Cry j j                | 3.20e-01 |
| 32 | 111 | 13.8 | 17  | 1 | R61580 | Cedar pollen allergen  | 5.22e-01 |
| 33 | 111 | 13.8 | 17  | 1 | W63445 | Sugi allergen protein  | 5.22e-01 |
| 34 | 110 | 13.6 | 13  | 1 | W63445 | Japanese cedar pollen  | 5.22e-01 |
| 35 | 107 | 13.3 | 15  | 1 | R75769 | Residues 66-80 of Cry  | 1.80e-00 |
| 36 | 107 | 13.3 | 15  | 1 | R75769 | Japan cedar pollen mat | 1.80e-00 |
| 37 | 107 | 13.3 | 47  | 1 | W63445 | Sugi allergen protein  | 1.80e-00 |
| 38 | 106 | 13.2 | 13  | 1 | 1579   | Japanese cedar pollen  | 1.80e-00 |
| 39 | 106 | 13.2 | 15  | 1 | 1579   | Residues 76-90 of Cry  | 1.80e-00 |
| 40 | 106 | 13.2 | 15  | 1 | 497    | Japan cedar pollen mat | 1.80e-00 |
| 41 | 106 | 13.2 | 20  | 1 | 549    | Cry j j pollen allerg  | 1.80e-00 |
| 42 | 106 | 13.2 | 20  | 1 | 401598 | Cry j j Japanese Cedar | 1.80e-00 |
| 43 | 106 | 13.2 | 24  | 1 | R41580 | Cry j j pollen allerg  | 1.29e-00 |
| 44 | 106 | 13.2 | 30  | 1 | W44582 | T-cell epitope peptide | 1.29e-00 |
| 45 | 103 | 12.8 | 15  | 1 | R97886 | Japan cedar pollen mat | 2.21e-00 |

ALIGNMENTS

RESULT 1  
ID W27370 standard: peptide: 105 AA  
AC W27370:  
DI 24-MAR-1998 (first entry)  
DE Multi-epitope peptide used as immunotherapeutic agent #2  
KW Multi-epitope peptide; immunotherapeutic agent; allergic disease  
XW T-cell epitope region; allergen; lymphocyte; immunoglobulin E  
OS Synthetic.  
PN W09732600-A1  
PO 12-SEP-1997  
PF 10-MAR-1997: J00740  
PR 10-MAR-1996: JP-080702.  
PA (MEIP) MELI MILK PRO CO LTD.  
PI Dairiki K, Iwara A, Kino K, Kame A, Sone T:  
DR WPI: 97-470495/43.  
PT Peptide immunotherapeutic agent to treat allergic diseases  
PT Contains multi-epitope peptide containing T cell epitope regions  
PS Claim 6: Page 31: 58pp; Jan. 1996.  
CC The present sequence represents a multi-epitope peptide which is used as  
CC a new immunotherapeutic agent. It comprises T cell epitope regions from 2  
CC or more different allergens preferably linked via an amino acid lysine  
CC dimer(s), where the T cell epitope regions have a positivity index  
CC greater than 100 as measured in a patient group responding to the  
CC allergen; have at least 75% reactivity with lymphocytes from patients  
CC responding to the allergen; and are not reactive with immunoglobulin E  
CC (IgE) antibodies from patients responsive to the allergen. The agent can  
CC be used to prevent and treat a wide variety of allergic diseases, e.g. by  
CC desensitisation. Side effects, e.g. those mediated by IgE, are reduced.  
SQ Sequence 105 AA:  
Query Match 100.00; Score 806; DB 1; Length 105;  
Best Local Similarity 100.00; Pred. No. 2.14e-66;  
Matches 105: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

|      |   |
|------|---|
| DB 1 | MKVTVAFNPGNRRRVFKRVSNI:HGRIDIFASKNFHLKNTIGTGRWKNKPNILW 50 |
| QY 1 | MKVTVAFNPGNRRRVFKRVSNI:HGRIDIFASKNFHLKNTIGTGRWKNKPNILW 50 |
| DB 6 | FAKLTGFTLMGRRLKMPMYIAG:FDGRVGGIIAAYONPASWK 105            |
| QY 6 | FAKLTGFTLMGRRLKMPMYIAG:FDGRVGGIIAAYONPASWK 105            |

RESULT 2  
ID W27371 standard: peptide: 134 AA.  
AC W27371:

24-MAR-1998 (first entry)  
 Multi-epitope peptide used as immunotherapeutic agent #1.  
 Multi-epitope peptide: immunotherapeutic agent; allergic diseases.  
 T-cell epitope region: allergen; lymphocyte; immunoglobulin E.  
 Synthesis:  
 WO9732609-A1  
 12-SEP-1997  
 10-MAR-1998  
 10-MAR-1998 JP-080702  
 (MEIP) MELJI MILK PROD CO LTD,  
 Dairiki, K. Iwama A, Kino K, Kame A, Some T.  
 WPI: 97-470495/43.  
 Peptide immunotherapeutic agent to treat allergic diseases -  
 contains multi-epitope peptide containing T cell epitope regions  
 from different allergens  
 Claim 6: Page 42, 43pp, Japanese  
 The present sequence represents a multi-epitope peptide which is used as  
 a new immunotherapeutic agent. It comprises T cell epitope regions from 2  
 or more different allergens (preferably linked via arginine or lysine  
 dimers), where the T cell epitope regions have a positivity index  
 greater than 1.0 as measured in a patient group responding to the  
 allergen; have at least 70% reactivity with lymphocytes from patients  
 responding to the allergen; and are not reactive with immunoglobulin E  
 (IgE) antibodies from patients responsive to the allergen. The agent can  
 be used to prevent and treat a wide variety of allergic diseases, e.g. by  
 desensitization. Side effects, e.g. those mediated by IgE, are reduced.  
 Sequence 134 AA:  
 Query Match 69.6% Score 561: DB 1: Length 134:  
 Best Local Similarity 85.7% Pred. No. 1,66e-42: Indels 13: Gaps 1:  
 Matches 90: Conservative 1: Mismatches 1:  
 1 MKVIVAFNFGPNRRVFVKRVSNIHGRIIEFASKNHLKNTIGTGR 60  
 I IIII IIII IIII IIII IIII IIII IIII IIII IIII  
 1 MKVIVAFNFGPNRRVFVKRVSNIHGRIIEFASKNHLKNTIGTGR 60  
 I IIII IIII IIII IIII IIII IIII IIII IIII IIII  
 61 TAKTGITIMGHPHWTISGNNKIKMPYIAGKTFGRRAE 105  
 I IIII IIII IIII IIII IIII IIII IIII IIII IIII  
 61 TAKTGITIMGHPHWTISGNNKIKMPYIAGKTFGRRAE 92  
 I IIII IIII IIII IIII IIII IIII IIII IIII IIII

24-MAR-1998 (first entry)  
 Multi-epitope peptide used as immunotherapeutic agent #1.  
 Multi-epitope peptide: immunotherapeutic agent; allergic diseases.  
 T-cell epitope region: allergen; lymphocyte; immunoglobulin E.  
 Synthesis:  
 WO9732609-A1  
 12-SEP-1997  
 10-MAR-1998  
 10-MAR-1998 JP-080702  
 (MEIP) MELJI MILK PROD CO LTD,  
 Dairiki, K. Iwama A, Kino K, Kame A, Some T.  
 WPI: 97-470495/43.  
 Peptide immunotherapeutic agent to treat allergic diseases -  
 contains multi-epitope peptide containing T cell epitope regions  
 from different allergens  
 Claim 6: Page 42, 43pp, Japanese  
 The present sequence represents a multi-epitope peptide which is used as  
 a new immunotherapeutic agent. It comprises T cell epitope regions from 2  
 or more different allergens (preferably linked via arginine or lysine  
 dimers), where the T cell epitope regions have a positivity index  
 greater than 1.0 as measured in a patient group responding to the  
 allergen; have at least 70% reactivity with lymphocytes from patients  
 responding to the allergen; and are not reactive with immunoglobulin E  
 (IgE) antibodies from patients responsive to the allergen. The agent can  
 be used to prevent and treat a wide variety of allergic diseases, e.g. by  
 desensitization. Side effects, e.g. those mediated by IgE, are reduced.  
 Sequence 134 AA:  
 Query Match 69.6% Score 561: DB 1: Length 134:  
 Best Local Similarity 85.7% Pred. No. 1,66e-42: Indels 13: Gaps 1:  
 Matches 90: Conservative 1: Mismatches 1:  
 1 MKVIVAFNFGPNRRVFVKRVSNIHGRIIEFASKNHLKNTIGTGR 60  
 I IIII IIII IIII IIII IIII IIII IIII IIII IIII  
 1 MKVIVAFNFGPNRRVFVKRVSNIHGRIIEFASKNHLKNTIGTGR 60  
 I IIII IIII IIII IIII IIII IIII IIII IIII IIII  
 61 TAKTGITIMGHPHWTISGNNKIKMPYIAGKTFGRRAE 105  
 I IIII IIII IIII IIII IIII IIII IIII IIII IIII  
 61 TAKTGITIMGHPHWTISGNNKIKMPYIAGKTFGRRAE 92  
 I IIII IIII IIII IIII IIII IIII IIII IIII IIII

Query Match 46.2% Score 372: DB 1: Length 80:  
 Best Local Similarity 95.2% Pred. No. 4.23e-06:  
 Matches 20: Conservative 1: Mismatches 1: Indels 0: Gaps 0:

Query Match 21.3% Score 172: DB 1: Length 514:  
 Best Local Similarity 95.2% Pred. No. 4.23e-06:  
 Matches 20: Conservative 1: Mismatches 1: Indels 0: Gaps 0:





```

RESULT: 15
ID: R45541 standard: Protein: 374 AA.
AC: R45541:
ET: 13-JUL-1994 (first entry)
DE: Cry 1 : pollen allergen.

```

```

Query Katch      17.98% Score 144 DB 17 Length 172.0
Best local Similarity 85.0% P-Id. No. 1.05e-03
Matches 17: Conservative 13 Mismatches 2: Indels

DB      97 MNKRLKMPWYIAGYKIFDGR 116
      1 :|||||:|||||
QY      70 MSRLKMPWYIAGYKIFDGR 89

Search completed: Mon Jun 19 15:56:13 2000
Job time : 11 secs.

```

\*\*\*\*\*  
WATERMAN  
\*\*\*\*\*  
(TM)

Release: J. A. John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1991-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MSearch\_PP Protein - protein database search, using Smith-Waterman algorithm

Run on: Mon Jun 19 15:59:53 2000; Master time 6.13 Seconds  
405,568 Million cell updates/sec

Tabular output not generated.

Title: US-09-142-524A-2  
Description: (1-105) from US09.42524A.pap  
Project Source: 806  
Sequence: 1 MKVIVAFNFGPNRRVFKS.....FGRRVDGIIAAYONPASWK 105

Scoring table:  
Gap 150  
Gap 11

Searched: 188963 seqs, 23686106 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries  
Maximum DB seq length 105

Database: a-geneseq15  
a-geneseqp

Statistics: Mean 29.704; Variance 127.355; scale 0.233

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description                    | Pred. No. |
|------------|-------|-------------|--------|-------|--------------------------------|-----------|
| 1          | 806   | 100.0       | 105    | 1     | W27370 Multi-epitope peptide   | 2.14e-66  |
| 2          | 732   | 46.2        | 80     | 1     | W27369 Multi-epitope peptide   | 2.88e-24  |
| 3          | 138   | 17.1        | 17     | 1     | R81581 Cedar pollen allergen   | 3.12e-03  |
| 4          | 137   | 17.0        | 18     | 1     | W80346 Sugi allergen protein   | 4.02e-03  |
| 5          | 131   | 16.3        | 17     | 1     | R81588 Cedar pollen allergen   | 1.25e-02  |
| 6          | 131   | 16.3        | 20     | 1     | W42171 T-cell epitope peptide  | 1.26e-02  |
| 7          | 124   | 15.4        | 20     | 1     | R45550 Cry j I pollen allergen | 4.72e-02  |
| 8          | 124   | 15.4        | 20     | 1     | R82499 Cry j I Japanese Cedar  | 4.72e-02  |
| 9          | 124   | 15.4        | 30     | 1     | R45579 Cry j I pollen allergen | 4.72e-02  |
| 10         | 123   | 15.3        | 47     | 1     | W80357 Sugi allergen protein   | 5.70e-02  |
| 11         | 117   | 14.5        | 14     | 1     | W12535 Japanese cedar pollen   | 1.74e-01  |
| 12         | 117   | 14.5        | 15     | 1     | R97887 Japan cedar pollen mat  | 1.74e-01  |
| 13         | 117   | 14.5        | 17     | 1     | W80347 Sugi allergen protein   | 1.74e-01  |
| 14         | 115   | 14.3        | 17     | 1     | R81582 Cedar pollen allergen   | 2.51e-01  |
| 15         | 111   | 13.8        | 17     | 1     | R81580 Cedar pollen allergen   | 5.22e-01  |
| 16         | 111   | 13.8        | 17     | 1     | W80345 Sugi allergen protein   | 5.22e-01  |
| 17         | 110   | 13.6        | 13     | 1     | W12536 Japanese cedar pollen   | 6.47e-01  |
| 18         | 107   | 13.3        | 15     | 1     | W57760 Residues 66-80 of Cry   | 1.08e-02  |
| 19         | 107   | 13.3        | 15     | 1     | R97884 Japan cedar pollen mat  | 1.08e-02  |
| 20         | 107   | 13.3        | 47     | 1     | W80356 Sugi allergen protein   | 1.58e-02  |
| 21         | 106   | 13.2        | 13     | 1     | W12542 Japanese cedar pollen   | 1.29e-02  |
| 22         | 106   | 13.2        | 15     | 1     | W57761 Residues 76-90 of Cry   | 1.29e-02  |

Note: Post-processor removed 17 summaries from list due to search parameters chosen.

ALIGNMENTS

RESULT 1  
ID W27370 standard; peptide: 105 AA.  
AC W27370:  
DE 24-MAR-1998 (first entry)  
DE Multi-epitope peptide used as immunotherapeutic agent #1.  
KW Multi-epitope peptide; immunotherapeutic agent; allergic disease.  
KW T-cell epitope region; allergen; lymphocyte; immunoglobulin E.  
OS Synthetic.  
PN W09732600-A1.  
PD 12-SEP-1997.  
PF 10-MAR-1997; J00740.  
PR 10-MAR-1996; JP-080702.  
PA (WEIP) MEIJI MILK PROD CO LTD.  
PI Dairiki K, Iwama A, Kino K, Kume A, Sone T.  
DR WPI: 97-470495/43.  
PT Peptide immunotherapeutic agent to treat allergic diseases.  
PT contains multi-epitope peptide containing T cell epitope regions from different allergens.  
PS Claim 6; Page 31; 58pp; Japanese.  
CC The present sequence represents a multi-epitope peptide which is used as a new immunotherapeutic agent. It comprises T cell epitope regions from 2 or more different allergens (preferably linked via arginine or lysine dimers), where the T cell epitope regions have a positivity index greater than 100 as measured in a patient group responding to the allergen; have at least 70% reactivity with lymphocytes from patients responding to the allergen; and are not reactive with immunoglobulin E (IgE) antibodies from patients responsive to the allergen. The agent can be used to prevent and treat a wide variety of allergic diseases, e.g. by desensitization. Side effects, e.g. those mediated by IgE, are reduced.  
SC Sequence 105 AA:

Query Match: 100.0%; Score 806; DB ID: Length 105;  
Best Local Similarity 100.0%; P-Ind. No. 2.14e-66;  
Matches 105; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 1 MKVIVAFNFGPNRRVFKS...VIFGRRVDGIIAAYONPASWK 105  
QY 1 MKVIVAFNFGPNRRVFKS...VIFGRRVDGIIAAYONPASWK 105

DB 61 FAKLIGFILMGRLKMPYAGKYKFDGRRVDGIIAAYONPASWK 105  
QY 61 FAKLIGFILMGRLKMPYAGKYKFDGRRVDGIIAAYONPASWK 105

RESULT 2  
ID W27369 standard; peptide: 80 AA.  
AC W27369:  
DE 24-MAR-1998 (first entry)  
DE Multi-epitope peptide used as immunotherapeutic agent #1.  
KW Multi-epitope peptide; immunotherapeutic agent; allergic disease.  
KW T-cell epitope region; allergen; lymphocyte; immunoglobulin E.  
OS Synthetic.  
PN W09732600-A1.  
PD 12-SEP-1997.  
PF 10-MAR-1997; J00740.  
PR 10-MAR-1996; JP-080702.  
PA (WEIP) MEIJI MILK PROD CO LTD.  
PI Dairiki K, Iwama A, Kino K, Kume A, Sone T.  
DR WPI: 97-470495/43.  
PT Peptide immunotherapeutic agent to treat allergic diseases.  
PT contains multi-epitope peptide containing T cell epitope regions

from different allergens  
 Claim 6: Page 31: 58pp: Japanese.  
 The present sequence represents a multi-epitope peptide which is used as a new immunotherapeutic agent. It comprises T cell epitope regions from 2 or more different allergens (preferably linked via alanine or lysine dimers), where the T cell epitope regions have a positivity index greater than 100 as measured in a patient group responding to the allergen; have at least 70% reactivity with lymphocytes from patients according to the allergen; and are not reactive with immunoglobulin E (IgE) antibodies from patients responsive to the allergen. The agent can be used to prevent and treat a wide variety of allergic diseases, e.g. by desensitisation. Side effects, e.g. those mediated by IgE, are reduced.  
 Sequence 88 AA:

Query Match 46.2% Score 372; DB 1; Length 80;  
 Best Local Similarity 100.0%; Pred. No. 2,096-24;  
 Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 1 MKVTVAFNFGNNRRVFKKNSVNIHRRIDIFASKNFHLOKNTIGTCRR 51  
 IIII IIII IIII IIII IIII IIII IIII IIII IIII  
 QY 1 MKVTVAFNFGNNRRVFKKNSVNIHRRIDIFASKNFHLOKNTIGTCRR 51

RESULT 3  
 ID R81581 standard; Peptide: 17 AA.

AC R81581:  
 DE 24-MAY-1996 (first entry)  
 CE Cedar pollen allergen peptide 9 (T-cell epitope).  
 KW Cedar; Cryptomeria japonica; pollen; allergen; immunoglobulin E;  
 KR IgE; T-cell; epitope; antibody; pollinosis; therapy; immunotherapy.  
 OS Synthetic.  
 PN EP-700429-A2.  
 PD 13-MAR-1996.  
 PF 08-SEP-1995; 306295.  
 PR 10-SEP-1994; JP-242137.  
 RR 14-JUL-1995; JP-200221.  
 RR 14-JUL-1995; JP-200204.  
 RA (HAYB.) HAYASHIBARA SEIBUTSU KAGAKU.  
 RI Hiko K., Saito S., Taniguchi Y.  
 DR WP: 96-140976/15.  
 PI New peptide(s) derived from cedar pollen allergens - activate  
 PT allergen-specific T-cells, but not allergen-specific IgE antibodies.  
 PT used for treating cedar pollinosis.  
 PS Claim 4, Page 28: 36pp: English.  
 CC Synthetic peptides based on portions of cedar pollen allergens A  
 CC (R81586) and B (R81587) were tested for their ability to activate  
 CC cedar allergen-specific T-cells, but not allergen-specific IgE  
 CC antibodies. 6 peptides (R81582-R81587) were identified as T-cell  
 CC epitopes. These peptides, plus subsequences (R81581-79) essential  
 CC for T-cell recognition, and homologous peptides (R81588-95) can  
 CC be used as immunotherapeutic agents to treat or prevent cedar  
 CC pollinosis, avoiding side-effects such as anaphylaxis.  
 SQ Sequence 17 AA:

Query Match 17.1% Score 138; DB 1; Length 17;  
 Best Local Similarity 100.0%; Pred. No. 3,926-03;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 1 NRWLQFAKLIGFTLMG 17  
 IIII IIII IIII IIII  
 QY 55 NRWLQFAKLIGFTLMG 71

RESULT 4  
 ID W80345 standard; peptide: 16 AA.

AC W80345:  
 DE 11-JAN-1999 (first entry)  
 CE Sugi allergen protein CryJ2 derived epitope for T cells.  
 KW T cell; epitope; sugi allergen protein CryJ2; CryJ2; treatment;  
 KR sugi; pollinosis; allergic reaction; pollen.  
 OS Synthetic.  
 PN J1035996-A.  
 PD 29-SEP-1998.

PF 22-DEC-1997; 353448.  
 PR 24-DEC-1996; JP-343441.  
 RA (HAYB.) HAYASHIBARA SEIBUTSU KAGAKU.  
 RI (SANY.) SANKYO CO. LTD.  
 DR WP: 96-577037/49.  
 PT A linked T cell epitope peptide - used for the treatment of  
 PT S-gli-pollinosis.  
 PS Claim 7: Page 18: 21pp: Japanese.  
 CC W80339-59 represent epitopes for T cells, derived from the sugi allergen  
 CC proteins CryJ1 (W80339-46, W80350-53 and W80356-58) and CryJ2 (W80355-49  
 CC and W80354-55). The peptides are useful for the treatment of  
 CC sugi-pollinosis, an allergic reaction of the body to pollen.  
 SQ Sequence 18 AA:

Query Match 17.0% Score 137; DB 1; Length 18;  
 Best Local Similarity 94.1%; Pred. No. 4,026-03;  
 Matches 16; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

DB 1 RINWQFAKLIGFTLMGK 17  
 IIII IIII IIII  
 QY 56 RINWQFAKLIGFTLMGR 72

RESULT 5  
 ID R81588 standard; Peptide: 17 AA.

AC R81588:  
 DE 24-MAY-1996 (first entry)  
 CE Cedar pollen allergen peptide 14 (homologue).  
 KW Cedar; Cryptomeria japonica; pollen; allergen; immunoglobulin E;  
 KR IgE; T-cell; epitope; antibody; pollinosis; therapy; immunotherapy.  
 OS Synthetic.  
 PN EP-700429-A2.  
 PD 13-MAR-1996.  
 PF 08-SEP-1995; 306295.  
 PR 10-SEP-1994; JP-242137.  
 RR 14-JUL-1995; JP-200221.  
 RR 14-JUL-1995; JP-200204.  
 RA (HAYB.) HAYASHIBARA SEIBUTSU KAGAKU.  
 RI Hiko K., Saito S., Taniguchi Y.  
 DR WP: 96-140976/15.  
 PI New peptide(s) derived from cedar pollen allergens - activate  
 PT allergen-specific T-cells, but not allergen-specific IgE antibodies.  
 PT used for treating cedar pollinosis.  
 PS Disclosure; Page 32: 36pp: English.  
 CC Synthetic peptides based on portions of cedar pollen allergens A  
 CC (R81586) and B (R81587) were tested for their ability to activate  
 CC cedar allergen-specific T-cells, but not allergen-specific IgE  
 CC antibodies. 6 peptides (R81582-R81587) were identified as T-cell  
 CC epitopes. These peptides, plus subsequences (R81581-79) essential  
 CC for T-cell recognition, and homologous peptides (R81588-95) can  
 CC be used as immunotherapeutic agents to treat or prevent cedar  
 CC pollinosis, avoiding side-effects such as anaphylaxis.  
 SQ Sequence 17 AA:

Query Match 16.3% Score 131; DB 1; Length 17;  
 Best Local Similarity 94.1%; Pred. No. 1,266-02;  
 Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

DB 1 NRWLQFAKLIGFTLMG 17  
 IIII IIII IIII IIII  
 QY 55 NRWLQFAKLIGFTLMG 71

RESULT 6  
 ID W42171 standard; peptide: 20 AA.

AC W42171:  
 DE 16-JUN-1998 (first entry)  
 CE T-cell epitope peptide 51 for Japanese cypress pollen antigen Chao2.  
 KW Japanese cypress pollen; antigen; T-cell epitope; Chao2;  
 KR diagnosis; allergy; spring tree pollen disease; pollinosis.  
 OS Cramaeyparis obtusa.  
 PN W09747648-A1.





DE Sugi allergen protein Cry11 derived epitope for T cells.  
 KW T cell epitope: sugi allergen proteins Cry11; Cry12: treatment:  
 KW sugi-pollinosis; allergic reaction; pollen.  
 PS Synthesis.

PN J0255198-A  
 PD 29-SEP-1998. 153448  
 PE 22-DEC-1997. JP-343441  
 PR 24-DEC-1996. JP-343441  
 PA (HAYB ) HAYASHIBARA SEIBUTSU KAGAKU.  
 PS (SANY ) SANKYO CO LTD.  
 WI: 98-57737/49.  
 PI A linked T cell epitope peptide - used for the treatment of:  
 CC sugi-pollinosis.  
 CC Claim 10: Page 5: 21pp. Japanese.  
 CC W6339-58 represent epitopes for T cells, derived from the sugi allergen  
 CC proteins Cry11 (W6339-44, W6339-51 and W6339-58) and Cry12 (W6339-45-49  
 CC and W6339-55). The peptides are useful for the treatment of:  
 CC sugi-pollinosis, an allergic reaction of the body to pollen.  
 CC Sequence 14 AA.

Query Match 15 38; Score 123; DB 1; Length 47;  
 Best Local Similarity 57 18; Pred. No. 5,70p-32;  
 Matches 27; Conservative 2; Mismatches 16; Indels 2; Gaps 2;

DB 1 MKTVAFNCFKLTSGKASSGNG:IAAYQNPASWFAKSNFHLOKNT 47  
 QY 1 MKTVAFNCFGPNRRVF:KRVSNVTHG-RRIDI-FASKNFHLOKNI 45

## RESULT 11

ID W12535 standard; peptide: 14 AA.

AC W12535  
 DT 30-APR-1997 (first entry)  
 DE Japanese cedar pollen allergen-derived peptide 4.  
 KW Japanese cedar; sugi; pollen; pollinosis; pollenosis; allergen;  
 KW allergy; T-cell epitope; TCR; T-cell receptor; activation;  
 KW immune tolerance; Cry11  
 PS Cryptomeria japonica  
 PS Synthesis.  
 PN J0833391-A  
 PD 17-DEC-1996.  
 PE 18-JUN-1995. 161448  
 PR 07-APR-1995. JP-082519  
 PA (HAYB ) HAYASHIBARA SEIBUTSU KAGAKU.  
 PS (SANY ) SANKYO CO LTD.  
 WI: 97-095487/09.

PI Peptide allergen derived from Japanese cedar pollen - causes T cell  
 PI response specific to cedar pollen, for treatment of pollinosis  
 PS Claim 5: Page 19. 21pp. Japanese.  
 CC The present sequence is one of 24 claimed peptides which were  
 CC synthesised based on Japanese cedar pollen sequences. This peptide  
 CC was shown to have Japanese cedar pollen antigen T cell epitope  
 CC activity by using T cells isolated from a cedar pollinosis patient.  
 CC The peptide produces little or no anaphylaxis.  
 CC Sequence 14 AA.

Query Match 14 58; Score 117; DB 1; Length 14;  
 Best Local Similarity 100.0%; Pred. No. 1,74e-01;  
 Matches 14; Conservative 0; Mismatches 0; Indels 2; Gaps 0;

DB 1 WLQFAKLTGFTLMG 14  
 QY 58 WLQFAKLTGFTLMG 71

## RESULT 12

ID R97887 standard; peptide: 15 AA.

AC R97887  
 DT 16-JUN-1996 (first entry)  
 DE Japan cedar pollen mature allergen Cry j II amino acids 81-95.  
 KW Allergen; epitope; overlapping peptide; Cry j II; cedar pollen;  
 KW Sugi pollinosis; diagnosis; treatment.  
 PS Cryptomeria japonica.

PN J08047392-A.  
 PD 20-FEB-1996.  
 PE 07-NOV-1994. 297840.  
 PR 05-NOV-1993. JP-276773.  
 PA (MEIP ) MEIJI MILK PROD CO LTD.  
 WI: 96-166249/17.  
 PI Japan cedar pollen allergen Cry j II epitope - comprises at least  
 PI part of specified 450 amino acid protein  
 PS Claim 6: Page 3. 17pp. Japanese  
 CC R97871-R97960 are overlapping peptides used for the epitope mutation  
 CC of the Japan cedar pollen allergen Cry j II. Cry j II and allergen  
 CC peptides of it are useful in the diagnosis, prevention and treatment  
 CC of Sugi pollinosis, the allergic reaction to Japan cedar pollen.  
 CC Significant regions of the allergen were identified using the  
 CC overlapping peptides of the full epitope derived from a Cry j II  
 CC antigen-specific T cell line. Amino acids 66-80 (R97884) and 162-206  
 CC (R978908) of the full mature 450 amino acid allergen are the most  
 CC allergenic of the 90 peptides tested.  
 CC Sequence 15 AA.

Query Match 14 58; Score 137; DB 1; Length 15;  
 Best Local Similarity 100.0%; Pred. No. 1,74e-01;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 1 NNRWLOFAKLTGFT 15  
 QY 54 NNRWLOFAKLTGFT 68

## RESULT 13

ID W80347 standard; peptide: 17 AA.

AC W80347  
 DT 11-JAN-1999 (first entry)  
 DE Sugi allergen protein Cry12 derived epitope for T cells.  
 KW T cell epitope: sugi allergen proteins Cry11; Cry12: treatment:  
 KW sugi-pollinosis; allergic reaction; pollen.  
 PS Synthesis.

PN J10259198-A.  
 PD 29-SEP-1998.  
 PE 22-DEC-1997. 353448.  
 PR 24-DEC-1996. JP-343441  
 PA (HAYB ) HAYASHIBARA SEIBUTSU KAGAKU  
 PS (SANY ) SANKYO CO LTD.  
 WI: 98-57737/49.

PI A linked T cell epitope peptide - used for the treatment of:  
 PI sugi-pollinosis.  
 PS Claim 7: Page 19. 21pp. Japanese.  
 CC W6339-58 represent epitopes for T cells, derived from the sugi allergen  
 CC proteins Cry11 (W6339-44, W6339-51 and W6339-58) and Cry12 (W6339-45-49  
 CC and W6339-55). The peptides are useful for the treatment of:  
 CC sugi-pollinosis, an allergic reaction of the body to pollen.  
 CC Sequence 17 AA.

Query Match 14 58; Score 117; DB 1; Length 17;  
 Best Local Similarity 100.0%; Pred. No. 1,74e-01;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 1 IDIFASKNFHLOKNTIG 17  
 QY 31 IDIFASKNFHLOKNTIG 47

## RESULT 14

ID R81582 standard; peptide: 17 AA.

AC R81582  
 DT 24-MAY-1996 (first entry)  
 DE Cedar pollen allergen peptide 10 (T-cell epitope).  
 KW Cedar; Cryptomeria japonica; pollen; allergen; immunoglobulin E;  
 KW IgE; T-cell epitope; antibody; pollinosis; therapy; immunotherapy.  
 OS Synthetic.  
 PN EP-700929-A2.  
 PD 13-MAR-1996.

```

EP 08-SEP-1995: 306295.
PE 10-SEP-1994: JP-242137.
PR 14-JUL-1995: JP-200221.
PP 14-JUL-1995: JP-200234.
RA (HAYR : HAYASHIBARA SEIBUTSU KAGAKU.
RI Hino K, Saito S, Taniguchi Y.
R1 WPI: 96-140476/15.
R2 New peptides(s) derived from cedar pollen allergens - activate
R3 allergen-specific T-cells, but not allergen-specific IgE antibodies.
R4 used for treating cedar pollinosis.
R5 Claim 4: Para 28: 36pp: English.
R6 Synthetic peptides based on portions of cedar pollen allergens A
R7 (R81586) and B (R81587) were tested for their ability to activate
R8 cedar allergen-specific T-cells, but not allergen-specific IgE
R9 antibodies. 6 peptides (R81580-R81585) were identified as T-cell
R10 epitopes. These peptides, plus subsequent peptides (R81586-96), essential
R11 for T-cell reconstitution, and homologous peptides (R81588-96) can
R12 be used as immunotherapeutic agents to treat or prevent cedar
R13 pollinosis, avoiding side-effects such as anaphylaxis.
R14 Sequence: 17 AA:
R15
R16
R17
R18
R19
R20
R21
R22
R23
R24
R25
R26
R27
R28
R29
R30
R31
R32
R33
R34
R35
R36
R37
R38
R39
R40
R41
R42
R43
R44
R45
R46
R47
R48
R49
R50
R51
R52
R53
R54
R55
R56
R57
R58
R59
R60
R61
R62
R63
R64
R65
R66
R67
R68
R69
R70
R71
R72
R73
R74
R75
R76
R77
R78
R79
R80
R81
R82
R83
R84
R85
R86
R87
R88
R89
R90
R91
R92
R93
R94
R95
R96
R97
R98
R99
R100
R101
R102
R103
R104
R105
R106
R107
R108
R109
R110
R111
R112
R113
R114
R115
R116
R117
R118
R119
R120
R121
R122
R123
R124
R125
R126
R127
R128
R129
R130
R131
R132
R133
R134
R135
R136
R137
R138
R139
R140
R141
R142
R143
R144
R145
R146
R147
R148
R149
R150
R151
R152
R153
R154
R155
R156
R157
R158
R159
R160
R161
R162
R163
R164
R165
R166
R167
R168
R169
R170
R171
R172
R173
R174
R175
R176
R177
R178
R179
R180
R181
R182
R183
R184
R185
R186
R187
R188
R189
R190
R191
R192
R193
R194
R195
R196
R197
R198
R199
R200
R201
R202
R203
R204
R205
R206
R207
R208
R209
R210
R211
R212
R213
R214
R215
R216
R217
R218
R219
R220
R221
R222
R223
R224
R225
R226
R227
R228
R229
R230
R231
R232
R233
R234
R235
R236
R237
R238
R239
R240
R241
R242
R243
R244
R245
R246
R247
R248
R249
R250
R251
R252
R253
R254
R255
R256
R257
R258
R259
R260
R261
R262
R263
R264
R265
R266
R267
R268
R269
R270
R271
R272
R273
R274
R275
R276
R277
R278
R279
R280
R281
R282
R283
R284
R285
R286
R287
R288
R289
R290
R291
R292
R293
R294
R295
R296
R297
R298
R299
R300
R301
R302
R303
R304
R305
R306
R307
R308
R309
R310
R311
R312
R313
R314
R315
R316
R317
R318
R319
R320
R321
R322
R323
R324
R325
R326
R327
R328
R329
R330
R331
R332
R333
R334
R335
R336
R337
R338
R339
R340
R341
R342
R343
R344
R345
R346
R347
R348
R349
R350
R351
R352
R353
R354
R355
R356
R357
R358
R359
R360
R361
R362
R363
R364
R365
R366
R367
R368
R369
R370
R371
R372
R373
R374
R375
R376
R377
R378
R379
R380
R381
R382
R383
R384
R385
R386
R387
R388
R389
R390
R391
R392
R393
R394
R395
R396
R397
R398
R399
R400
R401
R402
R403
R404
R405
R406
R407
R408
R409
R410
R411
R412
R413
R414
R415
R416
R417
R418
R419
R420
R421
R422
R423
R424
R425
R426
R427
R428
R429
R430
R431
R432
R433
R434
R435
R436
R437
R438
R439
R440
R441
R442
R443
R444
R445
R446
R447
R448
R449
R450
R451
R452
R453
R454
R455
R456
R457
R458
R459
R460
R461
R462
R463
R464
R465
R466
R467
R468
R469
R470
R471
R472
R473
R474
R475
R476
R477
R478
R479
R480
R481
R482
R483
R484
R485
R486
R487
R488
R489
R490
R491
R492
R493
R494
R495
R496
R497
R498
R499
R500
R501
R502
R503
R504
R505
R506
R507
R508
R509
R510
R511
R512
R513
R514
R515
R516
R517
R518
R519
R520
R521
R522
R523
R524
R525
R526
R527
R528
R529
R530
R531
R532
R533
R534
R535
R536
R537
R538
R539
R540
R541
R542
R543
R544
R545
R546
R547
R548
R549
R550
R551
R552
R553
R554
R555
R556
R557
R558
R559
R560
R561
R562
R563
R564
R565
R566
R567
R568
R569
R570
R571
R572
R573
R574
R575
R576
R577
R578
R579
R580
R581
R582
R583
R584
R585
R586
R587
R588
R589
R590
R591
R592
R593
R594
R595
R596
R597
R598
R599
R600
R601
R602
R603
R604
R605
R606
R607
R608
R609
R610
R611
R612
R613
R614
R615
R616
R617
R618
R619
R620
R621
R622
R623
R624
R625
R626
R627
R628
R629
R630
R631
R632
R633
R634
R635
R636
R637
R638
R639
R640
R641
R642
R643
R644
R645
R646
R647
R648
R649
R650
R651
R652
R653
R654
R655
R656
R657
R658
R659
R660
R661
R662
R663
R664
R665
R666
R667
R668
R669
R670
R671
R672
R673
R674
R675
R676
R677
R678
R679
R680
R681
R682
R683
R684
R685
R686
R687
R688
R689
R690
R691
R692
R693
R694
R695
R696
R697
R698
R699
R700
R701
R702
R703
R704
R705
R706
R707
R708
R709
R710
R711
R712
R713
R714
R715
R716
R717
R718
R719
R720
R721
R722
R723
R724
R725
R726
R727
R728
R729
R730
R731
R732
R733
R734
R735
R736
R737
R738
R739
R740
R741
R742
R743
R744
R745
R746
R747
R748
R749
R750
R751
R752
R753
R754
R755
R756
R757
R758
R759
R760
R761
R762
R763
R764
R765
R766
R767
R768
R769
R770
R771
R772
R773
R774
R775
R776
R777
R778
R779
R780
R781
R782
R783
R784
R785
R786
R787
R788
R789
R790
R791
R792
R793
R794
R795
R796
R797
R798
R799
R800
R801
R802
R803
R804
R805
R806
R807
R808
R809
R810
R811
R812
R813
R814
R815
R816
R817
R818
R819
R820
R821
R822
R823
R824
R825
R826
R827
R828
R829
R830
R831
R832
R833
R834
R835
R836
R837
R838
R839
R840
R841
R842
R843
R844
R845
R846
R847
R848
R849
R850
R851
R852
R853
R854
R855
R856
R857
R858
R859
R860
R861
R862
R863
R864
R865
R866
R867
R868
R869
R870
R871
R872
R873
R874
R875
R876
R877
R878
R879
R880
R881
R882
R883
R884
R885
R886
R887
R888
R889
R890
R891
R892
R893
R894
R895
R896
R897
R898
R899
R900
R901
R902
R903
R904
R905
R906
R907
R908
R909
R910
R911
R912
R913
R914
R915
R916
R917
R918
R919
R920
R921
R922
R923
R924
R925
R926
R927
R928
R929
R930
R931
R932
R933
R934
R935
R936
R937
R938
R939
R940
R941
R942
R943
R944
R945
R946
R947
R948
R949
R950
R951
R952
R953
R954
R955
R956
R957
R958
R959
R960
R961
R962
R963
R964
R965
R966
R967
R968
R969
R970
R971
R972
R973
R974
R975
R976
R977
R978
R979
R980
R981
R982
R983
R984
R985
R986
R987
R988
R989
R990
R991
R992
R993
R994
R995
R996
R997
R998
R999
R1000
R1001
R1002
R1003
R1004
R1005
R1006
R1007
R1008
R1009
R1010
R1011
R1012
R1013
R1014
R1015
R1016
R1017
R1018
R1019
R1020
R1021
R1022
R1023
R1024
R1025
R1026
R1027
R1028
R1029
R1030
R1031
R1032
R1033
R1034
R1035
R1036
R1037
R1038
R1039
R1040
R1041
R1042
R1043
R1044
R1045
R1046
R1047
R1048
R1049
R1050
R1051
R1052
R1053
R1054
R1055
R1056
R1057
R1058
R1059
R1060
R1061
R1062
R1063
R1064
R1065
R1066
R1067
R1068
R1069
R1070
R1071
R1072
R1073
R1074
R1075
R1076
R1077
R1078
R1079
R1080
R1081
R1082
R1083
R1084
R1085
R1086
R1087
R1088
R1089
R1090
R1091
R1092
R1093
R1094
R1095
R1096
R1097
R1098
R1099
R1100
R1101
R1102
R1103
R1104
R1105
R1106
R1107
R1108
R1109
R1110
R1111
R1112
R1113
R1114
R1115
R1116
R1117
R1118
R1119
R1120
R1121
R1122
R1123
R1124
R1125
R1126
R1127
R1128
R1129
R1130
R1131
R1132
R1133
R1134
R1135
R1136
R1137
R1138
R1139
R1140
R1141
R1142
R1143
R1144
R1145
R1146
R1147
R1148
R1149
R1150
R1151
R1152
R1153
R1154
R1155
R1156
R1157
R1158
R1159
R1160
R1161
R1162
R1163
R1164
R1165
R1166
R1167
R1168
R1169
R1170
R1171
R1172
R1173
R1174
R1175
R1176
R1177
R1178
R1179
R1180
R1181
R1182
R1183
R1184
R1185
R1186
R1187
R1188
R1189
R1190
R1191
R1192
R1193
R1194
R1195
R1196
R1197
R1198
R1199
R1200
R1201
R1202
R1203
R1204
R1205
R1206
R1207
R1208
R1209
R1210
R1211
R1212
R1213
R1214
R1215
R1216
R1217
R1218
R1219
R1220
R1221
R1222
R1223
R1224
R1225
R1226
R1227
R1228
R1229
R1230
R1231
R1232
R1233
R1234
R1235
R1236
R1237
R1238
R1239
R1240
R1241
R1242
R1243
R1244
R1245
R1246
R1247
R1248
R1249
R1250
R1251
R1252
R1253
R1254
R1255
R1256
R1257
R1258
R1259
R1260
R1261
R1262
R1263
R1264
R1265
R1266
R1267
R1268
R1269
R1270
R1271
R1272
R1273
R1274
R1275
R1276
R1277
R1278
R1279
R1280
R1281
R1282
R1283
R1284
R1285
R1286
R1287
R1288
R1289
R1290
R1291
R1292
R1293
R1294
R1295
R1296
R1297
R1298
R1299
R1300
R1301
R1302
R1303
R1304
R1305
R1306
R1307
R1308
R1309
R1310
R1311
R1312
R1313
R1314
R1315
R1316
R1317
R1318
R1319
R1320
R1321
R1322
R1323
R1324
R1325
R1326
R1327
R1328
R1329
R1330
R1331
R1332
R1333
R1334
R1335
R1336
R1337
R1338
R1339
R1340
R1341
R1342
R1343
R1344
R1345
R1346
R1347
R1348
R1349
R1350
R1351
R1352
R1353
R1354
R1355
R1356
R1357
R1358
R1359
R1360
R1361
R1362
R1363
R1364
R1365
R1366
R1367
R1368
R1369
R1370
R1371
R1372
R1373
R1374
R1375
R1376
R1377
R1378
R1379
R1380
R1381
R1382
R1383
R1384
R1385
R1386
R1387
R1388
R1389
R1390
R1391
R1392
R1393
R1394
R1395
R1396
R1397
R1398
R1399
R1400
R1401
R1402
R1403
R1404
R1405
R1406
R1407
R1408
R1409
R1410
R1411
R1412
R1413
R1414
R1415
R1416
R1417
R1418
R1419
R1420
R1421
R1422
R1423
R1424
R1425
R1426
R1427
R1428
R1429
R1430
R1431
R1432
R1433
R1434
R1435
R1436
R1437
R1438
R1439
R1440
R1441
R1442
R1443
R1444
R1445
R1446
R1447
R1448
R1449
R1450
R1451
R1452
R1453
R1454
R1455
R1456
R1457
R1458
R1459
R1460
R1461
R1462
R1463
R1464
R1465
R1466
R1467
R1468
R1469
R1470
R1471
R1472
R1473
R1474
R1475
R1476
R1477
R1478
R1479
R1480
R1481
R1482
R1483
R1484
R1485
R1486
R1487
R1488
R1489
R1490
R1491
R1492
R1493
R1494
R1495
R1496
R1497
R1498
R1499
R1500
R1501
R1502
R1503
R1504
R1505
R1506
R1507
R1508
R1509
R1510
R1511
R1512
R1513
R1514
R1515
R1516
R1517
R1518
R1519
R1520
R1521
R1522
R1523
R1524
R1525
R1526
R1527
R1528
R1529
R1530
R1531
R1532
R1533
R1534
R1535
R1536
R1537
R1538
R1539
R1540
R1541
R1542
R1543
R1544
R1545
R1546
R1547
R1548
R1549
R1550
R1551
R1552
R1553
R1554
R1555
R1556
R1557
R1558
R1559
R1560
R1561
R1562
R1563
R1564
R1565
R1566
R1567
R1568
R1569
R1570
R1571
R1572
R1573
R1574
R1575
R1576
R1577
R1578
R1579
R1580
R1581
R1582
R1583
R1584
R1585
R1586
R1587
R1588
R1589
R1590
R1591
R1592
R1593
R1594
R1595
R1596
R1597
R1598
R1599
R1600
R1601
R1602
R1603
R1604
R1605
R1606
R1607
R1608
R1609
R1610
R1611
R1612
R1613
R1614
R1615
R1616
R1617
R1618
R1619
R1620
R1621
R1622
R1623
R1624
R1625
R1626
R1627
R1628
R1629
R1630
R1631
R1632
R1633
R1634
R1635
R1636
R1637
R1638
R1639
R1640
R1641
R1642
R1643
R1644
R1645
R1646
R1647
R1648
R1649
R1650
R1651
R1652
R1653
R1654
R1655
R1656
R1657
R1658
R1659
R1660
R1661
R1662
R1663
R1664
R1665
R1666
R1667
R1668
R1669
R1670
R1671
R1672
R1673
R1674
R1675
R1676
R1677
R1678
R1679
R1680
R1681
R1682
R1683
R1684
R1685
R1686
R1687
R1688
R1689
R1690
R1691
R1692
R1693
R1694
R1695
R1696
R1697
R1698
R1699
R1700
R1701
R1702
R1703
R1704
R1705
R1706
R1707
R1708
R1709
R1710
R1711
R1712
R1713
R1714
R1715
R1716
R1717
R1718
R1719
R1720
R1721
R1722
R1723
R1724
R1725
R1726
R1727
R1728
R1729
R1730
R1731
R1732
R1733
R1734
R1735
R1736
R1737
R1738
R1739
R1740
R1741
R1742
R1743
R1744
R1745
R1746
R1747
R1748
R1749
R1750
R1751
R1752
R1753
R1754
R1755
R1756
R1757
R1758
R1759
R1760
R1761
R1762
R1763
R1764
R1765
R1766
R1767
R1768
R1769
R1770
R1771
R1772
R1773
R1774
R1775
R1776
R1777
R1778
R1779
R1780
R1781
R1782
R1783
R1784
R1785
R1786
R1787
R1788
R1789
R1790
R1791
R1792
R1793
R1794
R1795
R1796
R1797
R1798
R1799
R1800
R1801
R1802
R1803
R1804
R1805
R1806
R1807
R1808
R1809
R1810
R1811
R1812
R1813
R1814
R1815
R1816
R1817
R1818
R1819
R1820
R1821
R1822
R1823
R1824
R1825
R1826
R1827
R1828
R1829
R1830
R1831
R1832
R1833
R1834
R1835
R1836
R1837
R1838
R1839
R1840
R1841
R1842
R1843
R1844
R1845
R1846
R1847
R1848
R1849
R1850
R1851
R1852
R1853
R1854
R1855
R1856
R1857
R1858
R1859
R1860
R1861
R1862
R1863
R1864
R1865
R1866
R1867
R1868
R1869
R1870
R1871
R1872
R1873
R1874
R1875
R1876
R1877
R1878
R1879
R1880
R1881
R1882
R1883
R1884
R1885
R1886
R1887
R1888
R1889
R1890
R1891
R1892
R1893
R1894
R1895
R1896
R1897
R1898
R1899
R1900
R1901
R1902
R1903
R1904
R1905
R1906
R1907
R1908
R1909
R1910
R1911
R1912
R1913
R1914
R1915
R1916
R1917
R1918
R1919
R1920
R1921
R1922
R1923
R1924
R1925
R1926
R1927
R1928
R1929
R1930
R1931
R1932
R1933
R1934
R1935
R1936
R1937
R1938
R1939
R1940
R1941
R1942
R1943
R1944
R1945
R1946
R1947
R1948
R1949
R1950
R1951
R1952
R1953
R1954
R1955
R1956
R1957
R1958
R1959
R1960
R1961
R1962
R1963
R1964
R1965
R1966
R1967
R1968
R1969
R1970
R1971
R1972
R1973
R1974
R1975
R1976
R1977
R1978
R1979
R1980
R1981
R1982
R1983
R1984
R1985
R1986
R1987
R1988
R1989
R1990
R1991
R1992
R1993
R1994
R1995
R1996
R1997
R1998
R1999
R2000
R2001
R2002
R2003
R2004
R2005
R2006
R2007
R2008
R2009
R2010
R2011
R2012
R2013
R2014
R2015
R2016
R2017
R2018
R2019
R2020
R2021
R2022
R2023
R2024
R2025
R2026
R2027
R2028
R2029
R2030
R2031
R2032
R2033
R2034
R2035
R2036
R2037
R2038
R2039
R2040
R2041
R2042
R2043
R2044
R2045
R2046
R2047
R2048
R2049
R2050
R2051
R2052
R2053
R2054
R2055
R2056
R2057
R2058
R2059
R2060
R2061
R2062
R2063
R2064
R2065
R2066
R2067
R2068
R2069
R2070
R2071
R2072
R2073
R2074
R2075
R2076
R2077
R2078
R2079
R2080
R2081
R2082
R2083
R2084
R2085
R2086
R2087
R2088
R2089
R2090
R2091
R2092
R2093
R2094
R2095
R2096
R2097
R2098
R2099
R2100
R2101
R2102
R2103
R2104
R2105
R2106
R2107
R2108
R2109
R2110
R2111
R2112
R2113
R2114
R2115
R2116
R2117
R2118
R2119
R2120
R2121
R2122
R2123
R2124
R2125
R2126
R2127
R2128
R2129
R2130
R2131
R2132
R2133
R2134
R2135
R2136
R2137
R2138
R2139
R2140
R2141
R2142
R2143
R2144
R2145
R2146
R2147
R2148
R2149
R2150
R2151
R2152
R2153
R2154
R2155
R2156
R2157
R2158
R2159
R2160
R2161
R2162
R2163
R2164
R216
```

.....  
[M][E][S][S][A][G][E]  
..... (CM)  
.....

Release 3.1A John F. Collins, BioComputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPatch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Tue Jun 20 13:38:29 2000: MaxPar time 4.18 Seconds  
352.971 Million cell updates/sec

Tabular output not generated.

Title: US-09-142-524A-2  
Description: (1-105) from US09142524A.pep  
Perfect Score: 808  
Sequence: 1 MKVTVAENQFGPNRVEIKR.....FDGRVDGIIAAYQNFASNK 105

Scoring table: PAM 150  
Gap 11

Searched: 145341 seqs, 14437480 residues

Post-processing: Minimum Match 08  
Listing first 45 summaries  
Maximum DB seq length 105

Databases: 3-issued  
1:5A\_COMB 2:5B\_COMB 3:4\_COMB 4:PGT\_COMB 5:backfiles1

Statistics: Mean 27.828; Variance 123.399; scale 0.226

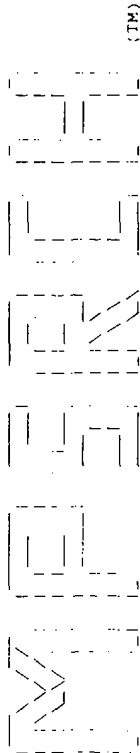
Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length DB | ID | Description | Pred. No. |
|------------|-------|-------------|-----------|----|-------------|-----------|
|------------|-------|-------------|-----------|----|-------------|-----------|

No matches found.

Search completed: Tue Jun 20 13:38:33 2000  
Job time : 5 secs.



\*\*\*\*\*  
Release 3.1A John F. Collins, BioComputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd  
\*\*\*\*\*  
Mitsch\_PP protein - protein database search, using Smith-Waterman algorithm  
Run on: Mon Jun 19 15:59:18 2000 Master time 9.36 Seconds  
529,224 Million cell updates/sec  
Tutular output not generated.  
Title: US-09-142-524A-2  
Description: (1-105) from US09142524A.ppt  
Perfect Score: 806  
Sequence: 1 MKVIVAFNFGPNRVRVTKR.....FDGRVDGIIAAYQNPSWK 105  
Scoring table: FAM 150  
Gap 11  
Searched: 14280 seqs, 47172406 residues  
Post-processing: Minimum Match 0%  
Listing first 45 summaries  
Maximum FR seq length 105  
Database: pir62  
pir61 2:pir2 3:pir3 4:pir4  
Statistics: Mean 41.762 Variance 84.448 scale 0.495  
Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed.  
and is derived by analysis of the total score distribution.  
Result Quality  
No. Score Match Length ID Description Prod. No.  
\*\*\*\*\*  
SUMMARIES  
\*\*\*\*\*  
No matches found.  
Search completed: Mon Jun 19 15:59:36 2000  
Job time : 18 secs.

(33)

20

mparch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Mon Jun 19 15:57:59 2000; MasPar time 6.72 Seconds

Tabular output not generated.

File: JS-09-142-524A-2

Description: (1-105) from US99142524A.ppt

Perfect Score: 896  
Sequence: 1 MKVTVAENCFGNRRVF~~KR~~.....ESGREVDGIIAAYQVFASWK 105

Source Lab.: PAM 150

Cap il

Searched: 83957 seqs, 30454973 residues

Post-processing: Minimum Match 08

Listing 1: First 45 summaries

Maximum DA seq length 105

Database: swiss-prot:38

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
84

Statistics: Mean 42.549; Variance 75.460; scale 0.565

pred. No. is the number of results predicted a chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

| Result | Score | Game Match | Length | Time | Prod. No.  |
|--------|-------|------------|--------|------|------------|
|        |       |            |        |      | Production |

No. matches found.

search completed: Mon Jun :9 15:58:19 2000

```

Job time : 21 secs.

```

.....  
MUSE (TM)  
.....

Release 3.1A John F. Collins, BioComputing Research Unit,  
Copyright (c) 1991-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

Musarchapp Protein - protein database search, using Smith-Waterman algorithm  
Run on: Mon Jun 19 15:58:35 2000 MasPar time 15.33 Seconds  
474.840 Million cell updates/sec  
Tabular output not generated.

Title: MUS-09-142-524A-2  
Description: (1:105) from USC9:42524A.pap  
Perfect Score: 806  
Sequence: 1 MKVTVAENGFGPNRRVFINK.....FQGRVVGIIAAYQNPASWK 105

Scoring table: PAM 150  
Gap 11

Searched: 225879 seqs, 69334122 residues

Post-processing: Minimum Match 08  
Listing first 45 summaries  
Maximum DB seq length 105

Database: spiremb12  
1:sp\_archaea 2:sp\_bacteria 3:sp\_fungal 4:sp\_human  
5:sp\_invertebrate 6:sp\_mammal 7:sp\_mhc 8:sp\_organelle  
9:sp\_phase 10:sp\_plant 11:sp\_rodent 12:sp\_unclassified  
13:sp\_vertebrate 14:sp\_virus

Statistics: Mean 41.638; Variance 72.759; Scale 0.573

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

| Result | Query | No. | Score | Match | Length | DB | ID | Description | Pred. No. |
|--------|-------|-----|-------|-------|--------|----|----|-------------|-----------|
|--------|-------|-----|-------|-------|--------|----|----|-------------|-----------|

No matches found.

Search completed: Mon Jun 19 15:59:02 2000  
Job time : 27 secs.

.....

WATERMAN

(TM)

.....

Release 1.1A John F. Collins, BioComputing Research Unit,  
Copyright (C) 1994-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

Search: protein protein database search, using Smith-Waterman algorithm  
Run On: Mon Jun 19 18:58:37 2000 MasPar time 6.77 Seconds  
Format output not deprecated. 468.570 Million cell updates/sec

Title: >US-09-142-524A-3  
Description: (1234) from US09142524A.pep  
RefSeq Score: 1026  
Sequence: 1 MKVTAFNFGNNRVFKR.....KFIKRVCGIIAAYONPASWK 134

Scoring table: PAM 150  
Gap 11

Search: 188463 seqs, 23686156 residues

Post processing: Minimum Match 08  
Listing first 45 summaries

Database: druckseq5  
Language: English

Statistics: Mean 30.692 Variance 132.267 scale 0.232

Pred. NO. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID       | Description             | Pred. No. |
|------------|-------|-------------|--------|----------|-------------------------|-----------|
| 1          | 1026  | 100.0       | 134    | 1 W27371 | Multi-epitope peptide   | 4 34e-87  |
| 2          | 561   | 54.7        | 205    | 1 W27372 | Multi-epitope peptide   | 6 58e-42  |
| 3          | 372   | 26.3        | 80     | 1 W27369 | Multi-epitope peptide   | 5 64e-24  |
| 4          | 293   | 26.3        | 353    | 1 R51388 | Japanese cedar pollen   | 6 30e-17  |
| 5          | 296   | 28.8        | 353    | 1 R51587 | Cedar pollen allergen   | 6 30e-17  |
| 6          | 296   | 28.8        | 374    | 1 R41937 | Cry j I                 | 6 30e-17  |
| 7          | 296   | 28.8        | 374    | 1 R61165 | Japanese cedar pollen   | 6 30e-17  |
| 8          | 296   | 28.8        | 374    | 1 R45541 | Cry j I pollen allergen | 6 30e-17  |
| 9          | 296   | 28.8        | 374    | 1 R82490 | Cry j I Japanese Cedar  | 6 30e-17  |
| 10         | 224   | 21.8        | 367    | 1 R45573 | Jun s I                 | 2 07e-10  |
| 11         | 224   | 21.8        | 370    | 1 R45578 | Jun v I                 | 2 07e-10  |
| 12         | 210   | 20.5        | 354    | 1 W42121 | Japanese cypress polle  | 3 61e-09  |
| 13         | 210   | 20.5        | 354    | 1 W43344 | Chamaecyparis obtusa p  | 3 61e-09  |
| 14         | 210   | 20.5        | 375    | 1 W43345 | Chamaecyparis obtusa p  | 3 61e-09  |
| 15         | 208   | 20.3        | 30     | 1 W46882 | T-cell epitope peptide  | 5 42e-09  |
| 16         | 172   | 16.8        | 460    | 1 R69731 | Japonicum allergen res  | 7 32e-06  |
| 17         | 172   | 16.8        | 514    | 1 R25385 | Cedar pollen allergen   | 7 32e-06  |
| 18         | 172   | 16.8        | 514    | 1 R31690 | Japanese cedar pollen   | 7 32e-06  |
| 19         | 172   | 16.8        | 514    | 1 R73599 | Japan cedar pollen Cry  | 7 32e-06  |
| 20         | 172   | 16.8        | 514    | 1 R73333 | Japanese cedar pollen   | 7 32e-06  |
| 21         | 172   | 16.8        | 514    | 1 R59792 | Japonicum allergen      | 7 32e-06  |
| 22         | 168   | 15.4        | 20     | 1 R45549 | Cry j I pollen allergen | 1 13e-04  |
| 23         | 158   | 15.4        | 20     | 1 R2498  | Cry j I Japanese Cedar  | 1 13e-04  |

|    |     |      |     |          |                         |          |
|----|-----|------|-----|----------|-------------------------|----------|
| 24 | 149 | 14.5 | 514 | 1 W43346 | Chamaecyparis obtusa p  | 6 45e-04 |
| 25 | 149 | 14.5 | 514 | 1 W42122 | Japanese cypress polle  | 6 45e-04 |
| 26 | 144 | 14.0 | 397 | 1 R43585 | Ragweed Pollen Allergen | 1 48e-03 |
| 27 | 142 | 13.8 | 30  | 1 R45579 | Cry j I pollen allergen | 2 41e-03 |
| 28 | 141 | 13.7 | 387 | 1 W41851 | Ragweed pollen Amb a 1  | 2 97e-03 |
| 29 | 141 | 13.7 | 387 | 1 R4333  | Amb a 1/Antigen E       | 2 97e-03 |
| 30 | 141 | 13.7 | 387 | 1 W4727  | Allergen Amb a 1A       | 2 97e-03 |
| 31 | 141 | 13.7 | 396 | 1 W2505  | Ragweed Amb a 1 allerg  | 2 97e-03 |
| 32 | 141 | 13.7 | 396 | 1 W4564  | Ragweed Pollen Allergen | 2 97e-03 |
| 33 | 141 | 13.7 | 397 | 1 R43245 | Amb a 1/Antigen E       | 2 97e-03 |
| 34 | 141 | 13.7 | 397 | 1 W58729 | Allergen Amb a 1C       | 2 97e-03 |
| 35 | 141 | 13.7 | 397 | 1 W62507 | Ragweed Amb a 1.3 aller | 2 97e-03 |
| 36 | 141 | 13.7 | 431 | 1 W41853 | Ragweed pollen Ant a 1  | 2 97e-03 |
| 37 | 138 | 13.5 | 17  | 1 R81581 | Cedar pollen allergen   | 5 24e-03 |
| 38 | 139 | 13.5 | 294 | 1 R07328 | Part of Amb a 1/Antige  | 4 34e-03 |
| 39 | 137 | 13.4 | 18  | 1 W03345 | Sugl allergen protein   | 6 44e-03 |
| 40 | 137 | 13.4 | 391 | 1 R07327 | Part of Amb a 1/Antige  | 6 44e-03 |
| 41 | 137 | 13.4 | 398 | 1 W58728 | Allergen Amb a 1A       | 6 43e-03 |
| 42 | 137 | 13.4 | 398 | 1 R43585 | Ragweed Pollen Allergen | 6 43e-03 |
| 43 | 137 | 13.4 | 398 | 1 R07334 | Amb a 1/Antigen E       | 6 43e-03 |
| 44 | 137 | 13.4 | 398 | 1 W2506  | Ragweed Amb a 1.2 aller | 6 43e-03 |
| 45 | 137 | 13.4 | 442 | 1 W4855  | Ragweed pollen UNC cla  | 6 33e-03 |

ALIGNMENTS

RESULT 1  
ID W27371 standard: peptide: 134 AA.  
AC W27371  
DI 24-MAR-1998 (first entry)  
DE Multi-epitope peptide used as immunotherapeutic agent  
KW Multi-epitope peptide; immunotherapeutic agent; allergic disease;  
KW T-cell epitope region; allergen; lymphocyte; immunoglobulin E;  
OS Synthetic  
PN W09732600.A1  
PD 12-SEP-1997  
PF 10-MAR-1997 JCO740  
PR 10-MAR-1996 JP-080702  
PA (MEIP) MELI MILK PROD CC LTD.  
PI Dairiki K, Iwama A, Kino K, Kame A, Sone T  
DR WPI: 97-470495/43  
PT Peptide immunotherapeutic agent to treat allergic diseases  
PT Contains multi-epitope peptide containing T cell epitope regions  
PT from different allergens  
PS Claim 5: Page 32: 58pp: Japanese  
CC The present sequence represents a multi-epitope peptide which is used as  
CC a new immunotherapeutic agent. It comprises 1 cell epitope regions from 2  
CC or more different allergens (preferably linked via an amino acid lysine  
CC dimers), where the T cell epitope regions have a positivity index  
CC greater than 100 as measured in a patient group responding to the  
CC allergen. have at least 70% reactivity with lymphocytes from patients  
CC responding to the allergen; and are not reactive with immunoglobulin E  
CC (IgE) antibodies from patients responsive to the allergen. The agent can  
CC be used to prevent and treat a wide variety of allergic diseases, e.g. by  
CC desensitisation. Side effects, e.g. those mediated by IgE, are reduced.  
SQ Sequence 134 AA:

|                       |  |                    |          |            |
|-----------------------|--|--------------------|----------|------------|
| Query Match           | 100.0%   | Score 1026         | DB 1     | Length 134 |
| Best Local Similarity | 100.0%   | Pred. NO. 4.34e-87 |          |            |
| Matches 134           | Conservative 0   | Mismatches 0       | Indels 0 | Gaps 0     |
| Db                    | 1 MKVTAFNFGNNRVFKRVSNIHGRRIDIFASKNFHLOKNTIGTRKWKNNRWLQ     | 60                 |          |            |
| QY                    | 1 MKVTAFNFGNNRVFKRVSNIHGRRIDIFASKNFHLOKNTIGTRKWKNNRWLQ     | 60                 |          |            |
| Db                    | 61 FAKLTGFTLMGRPRKLIIFSGNNMKLKMPLYIAGYKTFDGRRAEVSVHYNGAKFI | 120                |          |            |
| QY                    | 61 FAKLTGFTLMGRPRKLIIFSGNNMKLKMPLYIAGYKTFDGRRAEVSVHYNGAKFI | 120                |          |            |
| Db                    | 121 DGIIZAYONPASWK   | 134                |          |            |
| QY                    | 121 DGIIZAYONPASWK   | 134                |          |            |







```

CC can be used for the treatment and diagnosis of allergies associated
CC with Japanese cedar pollen.
SQ Sequence 367 AA:

Query Match                21.8% Score 224 DB 1 Length 957
Best Local Similarity 55.8% Pval N(2) 0.7e-10
Matches 25 Conservative 10 Miss-matches 13 Indels 5

DE 67 KALWIFSONMKIKLMKPYVAGSKILNIPGATVELUNGDPFSEKKAVSVY 100
      IIII I I II IIIII III III I IIII
CY 73 RLWLIFSOMMSIKLKNMYIAQKIFGGFEAEVSYYHVHNAKFIRWDVILL 124
      IIII I I IIIII III III I IIII

RESULT:
ID R45570 standard; Protein: 470 AA
AC R45570:
DI 11-JUL-1994 (first entry)
DE Sun v 1.
KW Japanese cedar; pollen allergen; allergy; treatment; diagnosis
KW T cell epitope; sensitivity; detection.
CS Juniperus virginiana. CS
FH Key Location/Qualifiers
FI peptide 1..21 "Signal peptide"
FI peptide 22..130c
FI /note= "Signal peptide"
FI /note= "Native peptide"
PD WC9401560-A.
PN 20-JAN-1994.
PR 15-JAN-1993: J00139.
PF 10-JUL-1992: WG-DC5661.
PK C1-SEP-1992: US-918992.
PR (IMMUNO) IMMUNOLOGIC PHARM CORP.
PA Board CF, Gorman RD, Griffith JJ, Kus M, Pollock JL
DR WPI: 94-C35066/C4
DR N-PSCB: Q55273.
PT Antigens derived from Japanese cedar pollen allergen Cry j 1
PT contain at least two T cell epitopes), used to treat or diagnose
PT allergy
PS Disclosure: Fig 17, 137pp; fish.
CC The sequence is that of Jun j 1, a homologue of the Japanese
CC cedar pollen allergen Cry j 1. Antigenic peptides derived from it
CC can be used for the treatment and diagnosis of allergies associated
CC with Japanese cedar pollen.
SQ Sequence 370 AA:

```

```

Query Match      21.8%  Score 224  DB 1  Length 370
Best Local Similarity 55.8%  Evalue NO. 2.07E-10
Matches 297  Conservative 191  Mismatches 13  Indels 1  Gaps 1

DB  57 KALWI:FSUNNN:KLMKPMYVVA:RSTLGGKSGALVYELNNOPTFLKPKYSHV 1-58
QY  73 RPIW:FSUNNN:KLMKPMYVVA:KIEDSPRAENVYVAVYAKPIFVYVGL 1-24

RESULT 12
ID  W42121 standard: protein. 354 AA.
AC  W42121:
DE  16-JUN-1998 (first entry)
DI  Japanese Cypress pollen antigen Chaol
KE  Japanese Cypress pollen antigen; 1-cell epitope; Chaol; Chaol2
KW  Japanese cypress pollen antigen; spring tree pollen disease; pollinosis
KW  Chamaecyparis; allergy; spring tree pollen disease; pollinosis
CS  Chamaecyparis obtusa.
PN  WO9747648-A1.
PC  18-DEC-1997.
PF  12-JUN-1997: J02031.
PF  14-JUN-1996: JP-153527.
PA  (WEIP) YELL; MILK PROD CO LTD.
PI  Dairiki K, Kino K.
DR  WPI: 98-05224/935.
DR  1-cell epitope peptide portion of Japanese cypress pollen antigens
P1  Chaol and Chaol2 - used for diagnosis and treatment of spring tree
P2  pollen disease
P3  Example 1: Page 14-15; 71pp; Japanese

```

CR N-PS23-13519  
 FI DNA encoding rhinocoryparis citrullus pollen allergen - T cell  
 PI epitope(s) of which are useful in development of preventative and  
 PT treating agent for Citrullus pollen allergic rhinitis

Search completed: Mon Jun 19 15:59:00 2000  
Job time : 23 secs.

.....  
M O E R E L E  
.....  
(TM)

.....  
M-Release 3.1A John F. Collins, Biocomputing Research Unit,  
Copyright (C) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd  
.....  
MSEARCH\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Tue Jun 20 13:33:35 2000: MasPar time 4.72 Seconds  
410.196 Million cell updates/sec  
Tabular output not generated.

Title: >US-09-142-524A-3  
Description: (1134) from US09142524A.ppt  
Perfect Score: 1026  
Sequence: 1 MKVTVPNQSPNRRVFIK.....KF:KRVIGI:AAVGNPMSK 134  
Scoring table: PAM 150  
Gap 11  
Searched: 145641 seqs, 14437480 residues  
Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: a-issued  
1:5A\_COMB 2:5B\_COMB 3:6\_COMB 4:PCI\_COMB 5:backfiles1  
Statistics: Mean 28.726; Variance 129.33; scale 0.222  
Pred No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed.  
and is derived by analysis of the total score distribution.

| Result |     | Query | Match | Length | DB         | ID          | Description | Pred. No. |
|--------|-----|-------|-------|--------|------------|-------------|-------------|-----------|
| 1      | 141 | 13.7  | 397   | 1      | US-08-175- | Sequence 72 | Applicati   | 2.06e-03  |
| 2      | 141 | 13.7  | 387   | 1      | US-08-232- | Sequence 72 | Applicati   | 2.06e-03  |
| 3      | 141 | 13.7  | 387   | 1      | US-08-250- | Sequence 72 | Applicati   | 2.06e-03  |
| 4      | 141 | 13.7  | 397   | 1      | US-08-175- | Sequence 76 | Applicati   | 2.06e-03  |
| 5      | 141 | 13.7  | 397   | 1      | US-08-230- | Sequence 76 | Applicati   | 2.06e-03  |
| 6      | 141 | 13.7  | 397   | 1      | US-08-230- | Sequence 76 | Applicati   | 2.06e-03  |
| 7      | 137 | 13.4  | 391   | 1      | US-08-175- | Sequence 59 | Applicati   | 4.31e-03  |
| 8      | 137 | 13.4  | 391   | 1      | US-08-230- | Sequence 59 | Applicati   | 4.31e-03  |
| 9      | 137 | 13.4  | 391   | 1      | US-08-232- | Sequence 59 | Applicati   | 4.31e-03  |
| 10     | 137 | 13.4  | 396   | 1      | US-08-175- | Sequence 74 | Applicati   | 4.31e-03  |
| 11     | 137 | 13.4  | 396   | 1      | US-08-230- | Sequence 74 | Applicati   | 4.31e-03  |
| 12     | 137 | 13.4  | 398   | 1      | US-08-230- | Sequence 74 | Applicati   | 4.31e-03  |
| 13     | 136 | 13.3  | 383   | 1      | US-08-175- | Sequence 78 | Applicati   | 5.18e-03  |
| 14     | 136 | 13.3  | 383   | 1      | US-08-230- | Sequence 78 | Applicati   | 5.18e-03  |
| 15     | 136 | 13.3  | 383   | 1      | US-08-230- | Sequence 78 | Applicati   | 5.18e-03  |
| 16     | 124 | 12.1  | 383   | 1      | US-08-175- | Sequence 80 | Applicati   | 4.62e-02  |
| 17     | 124 | 12.1  | 383   | 1      | US-08-230- | Sequence 80 | Applicati   | 4.62e-02  |
| 18     | 124 | 12.1  | 389   | 1      | US-08-230- | Sequence 80 | Applicati   | 4.62e-02  |
| 19     | 90  | 8.9   | 1079  | 3      | US-08-353- | Sequence 8  | Applicati   | 1.71e-01  |
| 20     | 90  | 8.8   | 1079  | 3      | US-08-485- | Sequence 8  | Applicati   | 1.71e-01  |
| 21     | 90  | 8.8   | 1079  | 2      | US-08-943- | Sequence 8  | Applicati   | 1.71e-01  |
| 22     | 90  | 8.8   | 1079  | 2      | US-08-480- | Sequence 8  | Applicati   | 1.71e-01  |
| 23     | 90  | 8.8   | 1079  | 1      | US-08-484- | Sequence 8  | Applicati   | 1.71e-01  |

|    |    |     |      |   |            |            |            |          |
|----|----|-----|------|---|------------|------------|------------|----------|
| 24 | 90 | 8.8 | 1085 | 3 | US-08-353- | Sequence 5 | Applicatio | 1.71e-01 |
| 25 | 90 | 8.8 | 1085 | 1 | US-08-485- | Sequence 5 | Applicatio | 1.71e-01 |
| 26 | 90 | 8.8 | 1085 | 2 | US-08-943- | Sequence 5 | Applicatio | 1.71e-01 |
| 27 | 90 | 8.8 | 1085 | 2 | US-08-485- | Sequence 5 | Applicatio | 1.71e-01 |
| 28 | 90 | 8.8 | 1085 | 1 | US-08-484- | Sequence 5 | Applicatio | 1.71e-01 |
| 29 | 89 | 8.7 | 1078 | 2 | US-08-484- | Sequence 7 | Applicatio | 2.02e-01 |
| 30 | 89 | 8.7 | 1078 | 3 | US-08-485- | Sequence 7 | Applicatio | 2.02e-01 |
| 31 | 89 | 8.7 | 1078 | 3 | US-08-353- | Sequence 7 | Applicatio | 2.02e-01 |
| 32 | 89 | 8.7 | 1078 | 1 | US-08-485- | Sequence 7 | Applicatio | 2.02e-01 |
| 33 | 89 | 8.7 | 1078 | 2 | US-08-943- | Sequence 7 | Applicatio | 2.02e-01 |
| 34 | 89 | 8.7 | 1088 | 3 | US-08-353- | Sequence 6 | Applicatio | 2.02e-01 |
| 35 | 89 | 8.7 | 1088 | 2 | US-08-943- | Sequence 6 | Applicatio | 2.02e-01 |
| 36 | 89 | 8.7 | 1088 | 1 | US-08-485- | Sequence 6 | Applicatio | 2.02e-01 |
| 37 | 89 | 8.7 | 1088 | 2 | US-08-484- | Sequence 6 | Applicatio | 2.02e-01 |
| 38 | 89 | 8.7 | 1088 | 2 | US-08-480- | Sequence 6 | Applicatio | 2.02e-01 |
| 39 | 89 | 8.7 | 1219 | 2 | US-08-687- | Sequence 6 | Applicatio | 2.02e-01 |
| 40 | 86 | 8.4 | 490  | 4 | PCI-US95-0 | Sequence 7 | Applicatio | 3.30e-01 |
| 41 | 86 | 8.4 | 490  | 1 | US-08-201- | Sequence 7 | Applicatio | 3.30e-01 |
| 42 | 86 | 8.4 | 490  | 2 | US-08-238- | Sequence 7 | Applicatio | 3.30e-01 |
| 43 | 86 | 8.4 | 802  | 2 | US-08-326- | Sequence 4 | Applicatio | 3.30e-01 |
| 44 | 86 | 8.4 | 802  | 1 | US-08-465- | Sequence 6 | Applicatio | 3.30e-01 |
| 45 | 86 | 8.4 | 802  | 1 | US-08-261- | Sequence 2 | Applicatio | 3.30e-01 |

ALIGNMENTS

RESULT 1  
ID US-08-175-069A-72 STANDARD: PRT: 367 AA.  
XX  
AC  
XX  
DT  
XX  
DE  
XX  
BE  
XX  
CC Sequence 72, Application US/08175069A  
CC Patent No. 577676;  
CC GENERAL INFORMATION:  
CC APPLICANT: Rogers, Bruce  
CC APPLICANT: Klapper, David G.  
CC APPLICANT: Rafnar, Thorunn  
CC APPLICANT: Kuo, Wei-chang  
CC TITLE OF INVENTION: Allergic Proteins From Recombinant US-  
CC NUMBER OF SEQUENCES: 9  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: LAHVE'S PATENT LLP  
CC STREET: 60 STATE STREET  
CC CITY: BOSTON  
CC STATE: MASSACHUSETTS  
CC COUNTRY: USA  
CC ZIP: 02109-1875  
CC COMPUTER READABLE FORM: disk  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: PatentIn Release #1.0, Version #1.25  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/175,069A  
CC FILING DATE: December 29, 1993  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 07/529,951  
CC FILING DATE: May 29, 1990  
CC APPLICATION NUMBER: US 07/325,365  
CC FILING DATE: March 17, 1989  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Amy E. Mandragoras  
CC REGISTRATION NUMBER: 36,207  
CC REFERENCE/DOCKET NUMBER: IMI-0180V  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (617)227-7400  
CC TELEFAX: (617)227-5941  
CC INFORMATION FOR SEQ ID NO: 72:  
CC SEQUENCE CHARACTERISTICS:



DE - Sequence 76, Application US/08175069A  
 XX  
 CC Sequence 76, Application US/08175069A  
 CC Patent No. 5776161  
 CC  
 CC GENERAL INFORMATION:  
 CC APPLICANT: Rogers, Bruce  
 CC APPLICANT: Klapper, David G.  
 CC APPLICANT: Rafnar, Thorunn  
 CC APPLICANT: Kuo, Mei-chang  
 CC TITLE OF INVENTION: Allergenic Proteins From Ragweed and Uses  
 CC NUMBER OF SEQUENCES: 93  
 CC CORRESPONDENCE ADDRESS:  
 CC ADDRESSEE: LAHIVE & COCKFIELD, LLP  
 CC STREET: 60 State Street  
 CC CITY: Boston  
 CC STATE: Massachusetts  
 CC COUNTRY: USA  
 CC ZIP: 02109-1875  
 CC COMPUTER READABLE FORM:  
 CC MEDIUM TYPE: Floppy disk  
 CC OPERATING SYSTEM: PC-DOS/MS-DOS  
 CC SOFTWARE: Patent in Release #1.0, Version #1.25  
 CC CURRENT APPLICATION DATA:  
 CC APPLICATION NUMBER: US/08/175-069A  
 CC FILING DATE: December 29, 1993  
 CC PRIOR APPLICATION DATA:  
 CC APPLICATION NUMBER: US 07/529,951  
 CC FILING DATE: May 29, 1990  
 CC ATTORNEY/AGENT INFORMATION:  
 CC NAME: Amy E. Macdragouras  
 CC REGISTRATION NUMBER: 36,207  
 CC REFERENCE/DOCKET NUMBER: IMI-018DV  
 CC TELECOMMUNICATION INFORMATION:  
 CC TELEPHONE: (617)227-7400  
 CC TELEFAX: (617)227-5941  
 CC INFORMATION FOR SEQ ID NO: 75:  
 CC SEQUENCE CHARACTERISTICS:  
 CC LENGTH: 397 amino acids  
 CC TYPE: amino acid  
 CC TOPOLOGY: linear  
 CC MOLECULE TYPE: protein  
 CC SEQUENCE 76, Application US/08290448A  
 XX  
 CC Query Match 13.7% Score 141 DB 1 Length 397:  
 CC Best Local Similarity 37.9% Pred. No. 2,06e-03:  
 CC Matches 22: Conservative 12: Mismatches 22: Indels 2: Gaps 2:  
 DE 108 FAAGNRPLWIFKNDMVIN ELVNSDKIIDRGVKE-I-INGGLIMNVKNII 163  
 QY 67 FTLMGRPLWIFSGNNIKLMPMYIAGYKTFDGRRAEVSVHVNGAKFKRYVGI 124

DE - Sequence 76, Application US/08290448A  
 XX  
 CC Sequence 76, Application US/08290448A  
 CC Patent No. 5698204  
 CC  
 CC GENERAL INFORMATION:  
 CC APPLICANT: Rogers, Bruce  
 CC APPLICANT: Klapper, David G.  
 CC APPLICANT: Rafnar, Thorunn  
 CC APPLICANT: Kuo, Mei-chang  
 CC TITLE OF INVENTION: Allergenic Proteins From Ragweed and Uses  
 CC NUMBER OF SEQUENCES: 93  
 CC CORRESPONDENCE ADDRESS:  
 CC ADDRESSEE: LAHIVE & COCKFIELD  
 CC STREET: 60 State Street, suite 510  
 CC CITY: Boston  
 CC STATE: Massachusetts  
 CC COUNTRY: USA  
 CC ZIP: 02109-1875  
 CC COMPUTER READABLE FORM:  
 CC MEDIUM TYPE: Floppy disk  
 CC COMPUTER: IBM PC compatible  
 CC OPERATING SYSTEM: PC-DOS/MS-DOS  
 CC SOFTWARE: Patent in Release #1.0, Version #1.25  
 CC CURRENT APPLICATION DATA:  
 CC APPLICATION NUMBER: US/08/290-448A  
 CC FILING DATE: August 15, 1994  
 CC PRIOR APPLICATION DATA:  
 CC APPLICATION NUMBER: US 07/529,951  
 CC FILING DATE: May 29, 1990  
 CC ATTORNEY/AGENT INFORMATION:  
 CC NAME: Amy E. Macdragouras  
 CC REGISTRATION NUMBER: 36,207  
 CC REFERENCE/DOCKET NUMBER: IMI-018CN  
 CC TELECOMMUNICATION INFORMATION:  
 CC TELEPHONE: (617)227-7400  
 CC TELEFAX: (617)227-5941  
 CC INFORMATION FOR SEQ ID NO: 76:  
 CC SEQUENCE CHARACTERISTICS:  
 CC LENGTH: 397 amino acids  
 CC TYPE: amino acid  
 CC TOPOLOGY: linear  
 CC MOLECULE TYPE: protein  
 CC SEQUENCE 397 AA: 42947 MW: 781425 CN:  
 SQ  
 CC Query Match 13.7% Score 141 DB 1 Length 397:  
 CC Best Local Similarity 37.9% Pred. No. 2,06e-03:  
 CC Matches 22: Conservative 12: Mismatches 22: Indels 2: Gaps 2:  
 DB 108 FAAGNRPLWIFKNDMVIN ELVNSDKIIDRGVKE-I-INGGLIMNVKNII 163  
 QY 67 FTLMGRPLWIFSGNNIKLMPMYIAGYKTFDGRRAEVSVHVNGAKFKRYVGI 124

CC SOFTWARE: Patent Release #1.0, Version #1.25  
 CC CURRENT APPLICATION DATA:  
 CC APPLICATION NUMBER: US/08/290,448A  
 CC FILING DATE: August 15, 1994  
 CC PRIOR APPLICATION DATA:  
 CC APPLICATION NUMBER: US 07/529,951  
 CC FILING DATE: May 29, 1990  
 CC APPLICATION NUMBER: US 07/325,365  
 CC FILING DATE: March 17, 1989  
 CC ATTORNEY/AGENT INFORMATION:  
 CC NAME: Amy E. Mandragoras  
 CC REGISTRATION NUMBER: 36,207  
 CC REFERENCE/DOCKET NUMBER: IMI-018CN  
 CC TELECOMMUNICATION INFORMATION:  
 CC TELEPHONE: (617)227-7400  
 CC TELEFAX: (617)227-7400  
 CC INFORMATION FOR SEQ ID NO: 163:  
 CC SEQUENCE CHARACTERISTICS:  
 CC LENGTH: 391 amino acids  
 CC TYPE: amino acid  
 CC TOPOLOGY: linear  
 CC MOLECULE TYPE: protein  
 CC SEQUENCE: 391 AA: 42864 MW: 761425.0N  
 CC  
 CC Query Match: 13.4% Score 141, DB 1, Length 197,  
 CC Best Local Similarity 37.9% Ref No. 436e-03,  
 CC Matches 22: Conservative 12: Mismatches 22: Indels 2: Gaps 2:  
 CC  
 CC 108 FAAGNRRPLMIIFKKNMVIHLNDELVNSKTIIDGRGVKVAIGILNNVKNI:163  
 CC  
 CC 67 FILMGRPLKLIIFSSNMNIKMKPMYIAGYKTFDSRAEVSIVHVGAKFIRVDGI:124  
 CC  
 CC  
 CC RESULT  
 CC TO US-08-075-869A-59 STANDARD: PRI: 391 AA:  
 CC AC xxxxxx  
 CC  
 CC Sequence 59, Application US/08-075-869A  
 CC Patent No. 5676364  
 CC GENERAL INFORMATION:  
 CC APPLICANT: Rogers, Bruce  
 CC APPLICANT: Klapper, David G.  
 CC APPLICANT: Rafsar, Thorun  
 CC APPLICANT: Kuo, Mel-Chand  
 CC TITLE OF INVENTION: Amino Acid Proteins From Bacteria and Fung  
 CC NUMBER OF SEQUENCES: 13  
 CC CORRESPONDENCE ADDRESS:  
 CC ADDRESSEE: LAHIVE & CUCKFIELD, LLP  
 CC STREET: 60 State Street  
 CC CITY: Boston  
 CC STATE: Massachusetts  
 CC COUNTRY: USA  
 CC ZIP: 02109-1875  
 CC COMPUTER READABLE FORM:  
 CC MEDIUM TYPE: Floppy disk  
 CC COMPUTER: IBM PC compatible  
 CC OPERATING SYSTEM: PC-DOS/MS-DOS  
 CC SOFTWARE: Patent Release #1.0, Version #1.25  
 CC CURRENT APPLICATION DATA:  
 CC APPLICATION NUMBER: US/08/290,448A  
 CC FILING DATE: August 15, 1994  
 CC PRIOR APPLICATION DATA:  
 CC APPLICATION NUMBER: US 07/529,951  
 CC FILING DATE: May 29, 1990  
 CC APPLICATION NUMBER: US 07/325,365  
 CC FILING DATE: March 17, 1989  
 CC ATTORNEY/AGENT INFORMATION:  
 CC NAME: Amy E. Mandragoras  
 CC REGISTRATION NUMBER: 36,207  
 CC REFERENCE/DOCKET NUMBER: IMI-018CN  
 CC TELECOMMUNICATION INFORMATION:  
 CC TELEPHONE: (617)227-7400  
 CC TELEFAX: (617)227-7400  
 CC INFORMATION FOR SEQ ID NO: 163:  
 CC SEQUENCE CHARACTERISTICS:  
 CC LENGTH: 391 amino acids  
 CC TYPE: amino acid  
 CC TOPOLOGY: linear  
 CC MOLECULE TYPE: protein

CC REGISTRATION NUMBER: 36,207  
 CC REFERENCE/DOCKET NUMBER: IMI-018CN  
 CC TELECOMMUNICATION INFORMATION:  
 CC TELEPHONE: (617)227-7400  
 CC TELEFAX: (617)227-7400  
 CC INFORMATION FOR SEQ ID NO: 59:  
 CC SEQUENCE CHARACTERISTICS:  
 CC LENGTH: 391 amino acid  
 CC TYPE: amino acid  
 CC TOPOLOGY: linear  
 CC MOLECULE TYPE: protein  
 CC SEQUENCE: 391 AA: 42864 MW: 761425.0N  
 CC  
 CC Query Match: 13.4% Score 141, DB 1, Length 197,  
 CC Best Local Similarity 39.7% Ref No. 436e-03,  
 CC Matches 23: Conservative 10: Mismatches 23: Indels 2: Gaps 2:  
 CC  
 CC DB 102 FAAGNRRPLMIIFKKNMVIHLNDELVNSKTIIDGRGVKVAIGILNNVKNI:163  
 CC  
 CC QY 67 FILMGRPLKLIIFSSNMNIKMKPMYIAGYKTFDSRAEVSIVHVGAKFIRVDGI:124  
 CC  
 CC  
 CC RESULT  
 CC TO US-08-290-448A-59 STANDARD: PRI: 391 AA:  
 CC AC xxxxxx  
 CC  
 CC Sequence 59, Application US/08-290-448A  
 CC Patent No. 5676354  
 CC GENERAL INFORMATION:  
 CC APPLICANT: Rogers, Bruce  
 CC APPLICANT: Klapper, David G.  
 CC APPLICANT: Rafsar, Thorun  
 CC APPLICANT: Kuo, Mel-Chand  
 CC TITLE OF INVENTION: Amino Acid Proteins From Bacteria and Fung  
 CC NUMBER OF SEQUENCES: 13  
 CC CORRESPONDENCE ADDRESS:  
 CC ADDRESSEE: LAHIVE & CUCKFIELD  
 CC STREET: 60 State Street, suite 510  
 CC CITY: Boston  
 CC STATE: Massachusetts  
 CC COUNTRY: USA  
 CC ZIP: 02109-1875  
 CC COMPUTER READABLE FORM:  
 CC MEDIUM TYPE: Floppy disk  
 CC COMPUTER: IBM PC compatible  
 CC OPERATING SYSTEM: PC-DOS/MS-DOS  
 CC SOFTWARE: Patent Release #1.0, Version #1.25  
 CC CURRENT APPLICATION DATA:  
 CC APPLICATION NUMBER: US/08/290,448A  
 CC FILING DATE: August 15, 1994  
 CC PRIOR APPLICATION DATA:  
 CC APPLICATION NUMBER: US 07/529,951  
 CC FILING DATE: May 29, 1990  
 CC APPLICATION NUMBER: US 07/325,365  
 CC FILING DATE: March 17, 1989  
 CC ATTORNEY/AGENT INFORMATION:  
 CC NAME: Amy E. Mandragoras  
 CC REGISTRATION NUMBER: 36,207  
 CC REFERENCE/DOCKET NUMBER: IMI-018CN  
 CC TELECOMMUNICATION INFORMATION:  
 CC TELEPHONE: (617)227-7400  
 CC TELEFAX: (617)227-7400  
 CC INFORMATION FOR SEQ ID NO: 59:  
 CC SEQUENCE CHARACTERISTICS:  
 CC LENGTH: 391 amino acids  
 CC TYPE: amino acid  
 CC TOPOLOGY: linear  
 CC MOLECULE TYPE: protein



CC Patent No. 5576934

```

0000 GENERAL INFORMATION:
0001 APPLICANT: Rodgers, Bruce
0002 APPLICANT: Klapper, David G.
0003 APPLICANT: Ratner, Thorunn
0004 APPLICANT: Kuo, Mei-chang
0005 TITLE OF INVENTION: Allergic Proteins From Ragweed and Uses
0006 NUMBER OF SEQUENCES: 93
0007 CORRESPONDENCE ADDRESS:
0008 ADDRESSEE: LAHIVE & COCKFIELD
0009 STREET: 60 State Street, suite 510
0010 CITY: Boston
0011 STATE: Massachusetts
0012 COUNTRY: USA
0013 ZIP: 02109-1875
0014 COMPUTER READABLE FORM:
0015 MEDIUM TYPE: Floppy disk
0016 COMPUTER: IBM PC compatible
0017 OPERATING SYSTEM: PC-DOS/MS-DOS
0018 SOFTWARE: Patman Release #1.0, Version #1.25
0019 CURRENT APPLICATION DATA:
0020 APPLICATION NUMBER: US/08/290,448A
0021 FILING DATE: August 15, 1994
0022 PRIOR APPLICATION DATA:
0023 APPLICATION NUMBER: US 07/529,957
0024 FILING DATE: May 29, 1990
0025 APPLICATION NUMBER: US 07/325,465
0026 FILING DATE: March 17, 1984
0027 ATTORNEY/AGENT INFORMATION:
0028 NAME: Amy E. Mandragoras
0029 REGISTRATION NUMBER: 35,207
0030 REFERENCE/DOCKET NUMBER: IM-018CN
0031 TELECOMMUNICATION INFORMATION:
0032 TELEPHONE: (617)227-7400
0033 TELEFAX: (617)227-5341
0034 INFORMATION FOR SEQ ID NO: 74:
0035 SEQUENCE CHARACTERISTICS:
0036 LENGTH: 399 amino acids
0037 TYPE: amino acid
0038 POPULATED: linear
0039 MOLECULE TYPE: protein
0040 SEQUENCE 399 AA: 43664 MW: 77306 CN:
0041 Query Match 13.4% Score 137: DB 1: Length 396;
0042 Best Loc: Similarity 39.7%; Pred.No.4.31e-03;
0043 Matches 23: Conservative 10: Mismatches 23: Gaps
0044
0045
0046
0047
0048
0049
0050
0051
0052
0053
0054
0055
0056
0057
0058
0059
0060
0061
0062
0063
0064
0065
0066
0067
0068
0069
0070
0071
0072
0073
0074
0075
0076
0077
0078
0079
0080
0081
0082
0083
0084
0085
0086
0087
0088
0089
0090
0091
0092
0093
0094
0095
0096
0097
0098
0099
0100
0101
0102
0103
0104
0105
0106
0107
0108
0109
0110
0111
0112
0113
0114
0115
0116
0117
0118
0119
0120
0121
0122
0123
0124
0125
0126
0127
0128
0129
0130
0131
0132
0133
0134
0135
0136
0137
0138
0139
0140
0141
0142
0143
0144
0145
0146
0147
0148
0149
0150
0151
0152
0153
0154
0155
0156
0157
0158
0159
0160
0161
0162
0163
0164
0165
0166
0167
0168
0169
0170
0171
0172
0173
0174
0175
0176
0177
0178
0179
0180
0181
0182
0183
0184
0185
0186
0187
0188
0189
0190
0191
0192
0193
0194
0195
0196
0197
0198
0199
0200
0201
0202
0203
0204
0205
0206
0207
0208
0209
0210
0211
0212
0213
0214
0215
0216
0217
0218
0219
0220
0221
0222
0223
0224
0225
0226
0227
0228
0229
0230
0231
0232
0233
0234
0235
0236
0237
0238
0239
0240
0241
0242
0243
0244
0245
0246
0247
0248
0249
0250
0251
0252
0253
0254
0255
0256
0257
0258
0259
0260
0261
0262
0263
0264
0265
0266
0267
0268
0269
0270
0271
0272
0273
0274
0275
0276
0277
0278
0279
0280
0281
0282
0283
0284
0285
0286
0287
0288
0289
0290
0291
0292
0293
0294
0295
0296
0297
0298
0299
0300
0301
0302
0303
0304
0305
0306
0307
0308
0309
0310
0311
0312
0313
0314
0315
0316
0317
0318
0319
0320
0321
0322
0323
0324
0325
0326
0327
0328
0329
0330
0331
0332
0333
0334
0335
0336
0337
0338
0339
0340
0341
0342
0343
0344
0345
0346
0347
0348
0349
0350
0351
0352
0353
0354
0355
0356
0357
0358
0359
0360
0361
0362
0363
0364
0365
0366
0367
0368
0369
0370
0371
0372
0373
0374
0375
0376
0377
0378
0379
0380
0381
0382
0383
0384
0385
0386
0387
0388
0389
0390
0391
0392
0393
0394
0395
0396
0397
0398
0399
0400
0401
0402
0403
0404
0405
0406
0407
0408
0409
0410
0411
0412
0413
0414
0415
0416
0417
0418
0419
0420
0421
0422
0423
0424
0425
0426
0427
0428
0429
0430
0431
0432
0433
0434
0435
0436
0437
0438
0439
0440
0441
0442
0443
0444
0445
0446
0447
0448
0449
0450
0451
0452
0453
0454
0455
0456
0457
0458
0459
0460
0461
0462
0463
0464
0465
0466
0467
0468
0469
0470
0471
0472
0473
0474
0475
0476
0477
0478
0479
0480
0481
0482
0483
0484
0485
0486
0487
0488
0489
0490
0491
0492
0493
0494
0495
0496
0497
0498
0499
0500
0501
0502
0503
0504
0505
0506
0507
0508
0509
0510
0511
0512
0513
0514
0515
0516
0517
0518
0519
0520
0521
0522
0523
0524
0525
0526
0527
0528
0529
0530
0531
0532
0533
0534
0535
0536
0537
0538
0539
0540
0541
0542
0543
0544
0545
0546
0547
0548
0549
0550
0551
0552
0553
0554
0555
0556
0557
0558
0559
0560
0561
0562
0563
0564
0565
0566
0567
0568
0569
0570
0571
0572
0573
0574
0575
0576
0577
0578
0579
0580
0581
0582
0583
0584
0585
0586
0587
0588
0589
0590
0591
0592
0593
0594
0595
0596
0597
0598
0599
0600
0601
0602
0603
0604
0605
0606
0607
0608
0609
0610
0611
0612
0613
0614
0615
0616
0617
0618
0619
0620
0621
0622
0623
0624
0625
0626
0627
0628
0629
0630
0631
0632
0633
0634
0635
0636
0637
0638
0639
0640
0641
0642
0643
0644
0645
0646
0647
0648
0649
0650
0651
0652
0653
0654
0655
0656
0657
0658
0659
0660
0661
0662
0663
0664
0665
0666
0667
0668
0669
0670
0671
0672
0673
0674
0675
0676
0677
0678
0679
0680
0681
0682
0683
0684
0685
0686
0687
0688
0689
0690
0691
0692
0693
0694
0695
0696
0697
0698
0699
0700
0701
0702
0703
0704
0705
0706
0707
0708
0709
0710
0711
0712
0713
0714
0715
0716
0717
0718
0719
0720
0721
0722
0723
0724
0725
```



US-09-142-524A-3.rai

| Hatches | 20 | Conservative                               | 10   | Mismatches | 24 | Indels | 20 | Gaps | 19 |
|---------|----|--|--|------------|----|--------|----|------|----|
| 20      | 99 | 60   | ANQNRNIIIFRINWVLDRELAINDKTGGGAKVEINAGFAYNWNI | 153        |    |        |    |      |    |
|         |    |  |  |            |    |        |    |      |    |
| 20      | 67 | FLMGREELIIFSSNNIKKRWPIYIACKTGGRAEVSIVHNGAF | 120  |            |    |        |    |      |    |

Search completed: 10 Jun 20 16:33:43 2000  
Job time : 8 secs.

.....  
[Logo]  
..... (TM)

Release J. A. John P. Collins, Biocomputing Research Unit,  
Copyright (c) 1999-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

Mpsrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Mon Jun 19 16:21:14 2000; MasPar time 29.67 seconds  
Tabular output not generated. 455,076 Million cell updates/sec

Title: US-09-142-524A-3  
Description: (1-134) from US09142524A.pap  
Perfect Score: 1026  
Sequence: 1 MKVVAFQFQPNRRVFIKRVNIIHGRIDIFASKNFHLQNTICTGRWKNNSINILQ 134

Scoring table: FAM 150  
Gap 11

Searches: 721208 seqs, 10076575 residues

Post processing: Minimum Match 0%

Listing first 45 summaries

Database: a-pending

Statistics: Mean 33.909; Variance 128.583; scale 0.264

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

| SUMMARY |     |       |                    | SUMMARY   |           |
|---------|-----|-------|--------------------|---|-----------|
| Result  | No. | Score | Match Length DB ID | Description                                     | Pred. No. |
| 1       | 126 | 100.0 | 134 17             | US-09-142- Sequence 3, Application US/09142524  | 1,61e-97  |
| 2       | 561 | 54.7  | 105 17             | US-08-468- Sequence 2, Application US/09142524  | 2,24e-46  |
| 3       | 372 | 36.3  | 80 17              | US-09-142- Sequence 1, Application US/09142524  | 3,77e-26  |
| 4       | 296 | 28.8  | 62 8               | US-08-350- Sequence 62, Application US/09142524 | 3,14e-18  |
| 5       | 296 | 28.8  | 60 7               | US-08-226- Sequence 62, Application US/09142524 | 3,14e-18  |
| 6       | 296 | 28.8  | 60 10              | US-08-467- Sequence 62, Application US/09142524 | 3,14e-18  |
| 7       | 296 | 28.8  | 60 10              | US-08-467- Sequence 62, Application US/09142524 | 3,14e-18  |
| 8       | 296 | 28.8  | 60 4               | US-07-938- Sequence 62, Application US/09142524 | 3,14e-18  |
| 9       | 296 | 28.8  | 60 10              | US-08-468- Sequence 62, Application US/09142524 | 3,14e-18  |
| 10      | 296 | 28.8  | 60 10              | US-08-467- Sequence 62, Application US/09142524 | 3,14e-18  |
| 11      | 296 | 28.8  | 353 11             | US-08-526- Sequence 15, Application US/09142524 | 3,14e-18  |
| 12      | 296 | 28.8  | 374 10             | US-08-467- Sequence 2, Application US/09142524  | 3,14e-18  |
| 13      | 296 | 28.8  | 374 10             | US-08-468- Sequence 2, Application US/09142524  | 3,14e-18  |
| 14      | 296 | 28.8  | 374 8              | US-07-730- Sequence 2, Application US/09142524  | 3,14e-18  |
| 15      | 296 | 28.8  | 374 8              | US-08-350- Sequence 2, Application US/09142524  | 3,14e-18  |
| 16      | 296 | 28.8  | 374 4              | US-07-938- Sequence 2, Application US/09142524  | 3,14e-18  |
| 17      | 296 | 28.8  | 374 7              | US-08-226- Sequence 2, Application US/09142524  | 3,14e-18  |
| 18      | 296 | 28.8  | 374 10             | US-08-467- Sequence 2, Application US/09142524  | 3,14e-18  |
| 19      | 296 | 28.8  | 374 10             | US-08-467- Sequence 2, Application US/09142524  | 3,14e-18  |
| 20      | 296 | 28.8  | 374 4              | US-07-729- Sequence 2, Application US/09142524  | 3,14e-18  |

|    |     |      |        |  |                                     |         |
|----|-----|------|--------|--|-------------------------------------|---------|
| 21 | 224 | 21.8 | 345 3  | US-09-142-524-3                                  | Sequence 1, Application US/09142524 | 5.1e-11 |
| 22 | 224 | 21.8 | 367 10 | US-08-350- Sequence 95, Application US/09142524  | 5.1e-11                             |         |
| 23 | 224 | 21.8 | 367 10 | US-08-468- Sequence 95, Application US/09142524  | 5.1e-11                             |         |
| 24 | 224 | 21.8 | 367 10 | US-08-467- Sequence 95, Application US/09142524  | 5.1e-11                             |         |
| 25 | 224 | 21.8 | 367 10 | US-08-467- Sequence 95, Application US/09142524  | 5.1e-11                             |         |
| 26 | 224 | 21.8 | 367 10 | US-08-467- Sequence 95, Application US/09142524  | 5.1e-11                             |         |
| 27 | 224 | 21.8 | 367 10 | US-08-467- Sequence 95, Application US/09142524  | 5.1e-11                             |         |
| 28 | 224 | 21.8 | 370 8  | US-08-467- Sequence 97, Application US/09142524  | 5.1e-11                             |         |
| 29 | 224 | 21.8 | 370 8  | US-08-467- Sequence 97, Application US/09142524  | 5.1e-11                             |         |
| 30 | 224 | 21.8 | 370 10 | US-08-467- Sequence 97, Application US/09142524  | 5.1e-11                             |         |
| 31 | 224 | 21.8 | 370 10 | US-08-467- Sequence 97, Application US/09142524  | 5.1e-11                             |         |
| 32 | 224 | 21.8 | 370 10 | US-08-467- Sequence 97, Application US/09142524  | 5.1e-11                             |         |
| 33 | 224 | 21.8 | 370 10 | US-08-468- Sequence 97, Application US/09142524  | 5.1e-11                             |         |
| 34 | 172 | 16.8 | 128 8  | US-08-450- Sequence 187, Application US/09142524 | 5.1e-06                             |         |
| 35 | 172 | 16.8 | 128 10 | US-08-468- Sequence 187, Application US/09142524 | 5.1e-06                             |         |
| 36 | 172 | 16.8 | 128 10 | US-08-467- Sequence 187, Application US/09142524 | 5.1e-06                             |         |
| 37 | 172 | 16.8 | 128 10 | US-08-467- Sequence 187, Application US/09142524 | 5.1e-06                             |         |
| 38 | 172 | 16.8 | 460 15 | US-08-980- Sequence 1, Application US/09142524   | 5.1e-06                             |         |
| 39 | 172 | 16.8 | 460 11 | US-08-571- Sequence 134, Application US/09142524 | 5.1e-06                             |         |
| 40 | 172 | 16.8 | 514 7  | US-09-225- Sequence 2, Application US/09142524   | 5.1e-06                             |         |
| 41 | 172 | 16.8 | 514 15 | US-08-980- Sequence 2, Application US/09142524   | 5.1e-06                             |         |
| 42 | 172 | 16.8 | 514 8  | US-08-334- Sequence 2, Application US/09142524   | 5.1e-06                             |         |
| 43 | 172 | 16.8 | 514 4  | US-07-976- Sequence 2, Application US/09142524   | 5.1e-06                             |         |
| 44 | 172 | 16.8 | 514 8  | US-08-350- Sequence 134, Application US/09142524 | 5.1e-06                             |         |
| 45 | 172 | 16.8 | 514 10 | US-08-468- Sequence 134, Application US/09142524 | 5.1e-06                             |         |

ALIGNMENTS

RESULT: 1  
ID US-09-142-524-3 STANDARD PRT: 134 AA.  
XX  
AC xxxxxx  
DT  
XX  
XX  
XX

DE Sequence 3, Application US/09142524

XX Sequence 3, Application US/09142524  
CC GENERAL INFORMATION:  
CC APPLICANT: Sore, Toshio  
CC APPLICANT: Kume, Akio  
CC APPLICANT: Kairiki, Kazuo  
CC APPLICANT: Iwama, Akiko  
CC APPLICANT: Kiro, Kousuke  
CC TITLE OF INVENTION: Peptide-based immunotherapeutic agent for treating  
CC TITLE OF INVENTION: Allergic diseases  
CC FILE REFERENCE: Docket No. SP-153  
CC CURRENT APPLICATION NUMBER: US/09142524  
CC CURRENT FILING DATE: 1999-03-10  
CC EARLIER APPLICATION NUMBER: B/89/702  
CC EARLIER FILING DATE: 1995-03-10  
CC EARLIER APPLICATION NUMBER: PCT/JP97/00740  
CC EARLIER FILING DATE: 1997-03-10  
CC NUMBER OF SEQ ID NOS: 5  
CC SOFTWARE: Patent In Ver. 2.0  
CC SEQ ID NO 3  
CC LENGTH: 134  
CC TYPE: PRT  
CC ORGANISM: Unknown  
CC FEATURE:  
CC OTHER INFORMATION: Description of Unknown Organism: peptide  
SQ SEQUENCE 134 AA: 15646 MW: 90017 CN:

Query Match 100.0%; Score 1026; DB 17; Length 134;  
Best Local Similarity 100.0%; Pred. No. 1,61e-97;  
Matches 134; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 MKVVAFQFQPNRRVFIKRVNIIHGRIDIFASKNFHLQNTICTGRWKNNSINILQ 60  
QY 1 MKVVAFQFQPNRRVFIKRVNIIHGRIDIFASKNFHLQNTICTGRWKNNSINILQ 60

DB 61 FAKLTGETLMGRRPLWIFSGNMNKLKMPYIAGYKTFDGRRAEVSYVHVGAKFISV 120



```
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (617) 466-6000  
CC TELEFAX: (617) 466-6040  
CC INFORMATION FOR SEQ ID NO: 62:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 60 amino acids  
CC TYPE: amine acid  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: peptide  
CC FRAGMENT TYPE: internal;  
SQ SEQUENCE 60 AA: 6544 MW: 11.64 kD:  
  
Query Match 28.8% Score 276; DB 7; Length 60  
Best Local Similarity 73.6% Pred Near 4,14e-16  
Matches 39; Conservative 5; Mismatches 7; Indels 2  
  
Db 6 RPLWIFSGNNIKLKMFMYIAGIKTIGRGAQV-YANGGPDVFIRKRVNVI 67  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
QY 73 RPLWIFSGNNIKLKMFMYIAGIKTIGRGAQV-YANGGPDVFIRKRVNVI 124  
  
RESULT 6 STANDARD ERT: 60 AA.  
ID US-08-467-006-62  
XX xxxxxx  
AC  
AD  
DT  
XX  
DE Sequence 62, Application US/ 657004  
XX  
XX Sequence 62, Application US/ 6457006  
CC GENERAL INFORMATION:  
CC APPLICANT: Griffith, Irwin J.  
CC APPLICANT: Pollock, Joanne  
CC APPLICANT: Bond, Julian  
CC APPLICANT: Garman, Richard D.  
CC APPLICANT: Kuo, Mei-Chang  
CC APPLICANT: Yeung, Si-Meng H.  
CC APPLICANT: Brauer, Andrew  
CC APPLICANT: Exley, Mark A.  
CC APPLICANT: Powers, Steven P.  
CC TITLE OF INVENTION: Allergic Proteins And Peptides From  
CC TITLE OF INVENTION: Japanese Cedar Pollen  
CC NUMBER OF SEQUENCES: 261  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Immunologic Pharmaceutical Corporation, Inc.  
CC STREET: 610 Lincoln St  
CC CITY: Waltham  
CC STATE: MA  
CC COUNTRY: USA  
CC ZIP: 02154  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: Patent In Release #1.0, Version #1.25  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/467,006  
CC FILING DATE: June 5, 1995  
CC CLASSIFICATION: 424  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: 08/350,225  
CC FILING DATE: December 6, 1994  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Jane E. Remillard  
CC REGISTRATION NUMBER: 38,872  
CC REFERENCE/DOCKET NUMBER: 025.6 USD6 (IMI-028CPD6)  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (617) 227-7400  
CC TELEFAX: (617) 227-5941  
CC INFORMATION FOR SEQ ID NO: 62:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 60 amino acids
```





```

RE-ULT 9
ID US-08-468-940-62 STANDARD: PRT: 60 AA.
XX
AC
XX
XX
DT
XX
DE
XX
Sequence 62, Application US/08468940
Sequence 62, Application US/08468940
GENERAL INFORMATION:
APPLICANT: Griffith, Irwin J.
APPLICANT: Pollock, Joanne
APPLICANT: Bond, Julian F.
APPLICANT: Gorman, Richard D.
APPLICANT: Kuo, Mei-Chang
APPLICANT: Yeung, Siu-mei H.
APPLICANT: Brauer, Andrew
APPLICANT: Exley, Mark A.
APPLICANT: Powers, Steven P.
TITLE OF INVENTION: Allergenic Proteins And Peptides From
TITLE OF INVENTION: Japanese Cedar Pollen
NUMBER OF SEQUENCES: 261
CORRESPONDENCE ADDRESS:
ADDRESS: Immunologic Pharmaceutical Corporation, Inc.
STREET: 610 Lincoln St
CITY: Waltham
STATE: MA
COUNTRY: USA
ZIP: 02154
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/468,940
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/450,225
FILING DATE: December 6, 1994
APPLICATION NUMBER: 08/226,248
FILING DATE: April 8, 1994
APPLICATION NUMBER: 07/938,990
FILING DATE: September 1, 1992
APPLICATION NUMBER: PC/US93/00139
FILING DATE: January 15, 1993
ATTORNEY/AGENT INFORMATION:
NAME: Parlese A. Vaccaro
REGISTRATION NUMBER: 35,724
REFERENCE/DOCKET NUMBER: 025.6 US (IM-028CP2)
TELEPHONE: (617) 466-6000
TELEFAX: (617) 466-6040
INFORMATION FOR SEQ ID NO: 62:
SEQUENCE CHARACTERISTICS:
LENGTH: 60 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
FRAGMENT TYPE: internal
SEQUENCE 50 AA: 6644 MW: 19464 CN:
Query Match 28.8% Score 296: DB 10: Length 60:
Best Local Similarity 73.6% Pred. No. 3,14e-18;
Matches 39: Conservative 5: Mismatches 7: Indels 2: Gaps 2:
DB 6 RPLWIFSGNNIKLKMPYIAGYKTFDGRGACV-YINGGPGCVFKRVSNVI 57
QY 73 RPLWIFSGNNIKLKMPYIAGYKTFDGRGACV-YINGGPGCVFKRVSNVI 124

```

```

RE-ULT 10
ID US-08-467-023-62 STANDARD: PRT: 60 AA.
XX
AC
XX
XX
DT
XX
DE
XX
Sequence 62, Application US/08467023
Sequence 62, Application US/08467023
GENERAL INFORMATION:
APPLICANT: Griffith, Irwin J.
APPLICANT: Pollock, Joanne
APPLICANT: Bond, Julian F.
APPLICANT: Gorman, Richard D.
APPLICANT: Kuo, Mei-Chang
APPLICANT: Yeung, Siu-mei H.
APPLICANT: Brauer, Andrew
APPLICANT: Exley, Mark A.
APPLICANT: Powers, Steven P.
TITLE OF INVENTION: Allergenic Proteins And Peptides From
TITLE OF INVENTION: Japanese Cedar Pollen
NUMBER OF SEQUENCES: 261
CORRESPONDENCE ADDRESS:
ADDRESS: Immunologic Pharmaceutical Corporation, Inc.
STREET: 610 Lincoln St
CITY: Waltham
STATE: MA
COUNTRY: USA
ZIP: 02154
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/467,023
FILING DATE: June 6, 1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/450,225
FILING DATE: December 6, 1994
ATTORNEY/AGENT INFORMATION:
NAME: Jane E. Remillard
REGISTRATION NUMBER: 38,872
REFERENCE/DOCKET NUMBER: 025.6 USC2 (IM-028CPD2)
TELEPHONE: (617) 227-7400
TELEFAX: (617) 227-7400
INFORMATION FOR SEQ ID NO: 62:
SEQUENCE CHARACTERISTICS:
LENGTH: 60 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
FRAGMENT TYPE: internal
SEQUENCE 50 AA: 6644 MW: 19464 CN:
Query Match 28.8% Score 296: DB 10: Length 60:
Best Local Similarity 73.6% Pred. No. 3,14e-18;
Matches 39: Conservative 5: Mismatches 7: Indels 2: Gaps 2:
DB 6 RPLWIFSGNNIKLKMPYIAGYKTFDGRGACV-YINGGPGCVFKRVSNVI 57
QY 73 RPLWIFSGNNIKLKMPYIAGYKTFDGRGACV-YINGGPGCVFKRVSNVI 124

```

XX Sequence 15, Application: US/08526179  
 XX Sequence 15, Application: US/08526179  
 CC GENERAL INFORMATION:  
 CC APPLICANT: Saito, Saburo  
 CC APPLICANT: Hino, Katsuhiko  
 CC APPLICANT: Taniuchi, Yoshitami  
 CC APPLICANT: Kuremoto, Katsuhiko  
 CC APPLICANT: Kuremoto, Katsuhiko  
 CC TITLE OF INVENTION: PEPTIDE AND USES OF THE SAME  
 CC NUMBER OF SEQUENCES: 24  
 CC CORRESPONDENCE ADDRESS:  
 CC ADDRESSEE: BROWN AND NEIMARK  
 CC STREET: 419 Seventh Street, N.W., Suite 100  
 CC CITY: Washington  
 CC STATE: D.C.  
 CC COUNTRY: USA  
 CC ZIP: 20004  
 CC COMPUTER READABLE FORM:  
 CC MEDIUM TYPE: Floppy disk  
 CC COMPUTER: IBM PC compatible  
 CC OPERATING SYSTEM: PC-DOS/MS-DOS  
 CC SOFTWARE: Patent Release #1.0, Version #1.30  
 CC CURRENT APPLICATION DATA:  
 CC APPLICATION NUMBER: US/08526179  
 CC FILING DATE:  
 CC CLASSIFICATION: 510  
 CC PRIORITY APPLICATION DATA:  
 CC APPLICATION NUMBER: JP 242,157/1994  
 CC FILING DATE: 10-SEP-1994  
 CC PRIORITY APPLICATION DATA:  
 CC APPLICATION NUMBER: JP 200,204/1995  
 CC FILING DATE: 14-JUL-1995  
 CC PRIORITY APPLICATION DATA:  
 CC APPLICATION NUMBER: JP 200,221/1995  
 CC FILING DATE: 14-JUL-1995  
 CC ATTORNEY/AGENT INFORMATION:  
 CC NAME: BROCKOFF, ROGER L.  
 CC REGISTRATION NUMBER: 25,618  
 CC REFERENCE/DOCKET NUMBER: SAITC-19  
 CC TELECOMMUNICATION INFORMATION:  
 CC TELEPHONE: 202-624-5197  
 CC TELEFAX: 202-737-3528  
 CC TELEX: 248633  
 CC INFORMATION FOR SEQ ID NO: 15:  
 CC SEQUENCE CHARACTERISTICS:  
 CC LENGTH: 353 amino acids  
 CC TYPE: amino acid  
 CC STRANDEDNESS: single  
 CC TOPOLOGY: linear  
 CC MOLECULE TYPE: Peptide  
 CC SEQUENCE 374 AA: 480-2 MW: 675794.70  
 CC Query Match 28.8% Score 2961 DB 11 Length 353  
 CC Best Local Similarity 77.6% Pct. Id. 34.0e-18  
 CC Matches 39: Conservative 7: Indels 2: Gaps 2:  
 CC 66 RPLWTFSSNNIKKPKMYIAGYKTFDGRGAV-YIGNSGPCVF\*KRVSNI 117  
 CC 73 RPLWTFSSNNIKKPKMYIAGYKTFDGRRAEVSYSVHNGA-KFIRRVGGII 124  
 CC RESULT 14  
 CC ID US-08-467-023-2 STANDARD: PRT: 374 AA.  
 CC XX xxxxxx  
 CC AC  
 CC DT  
 CC DE  
 CC Sequence 2, Application: US/08467023  
 CC Sequence 2, Application: US/08467023  
 CC GENERAL INFORMATION:  
 CC APPLICANT: Griffith, Irwin J.  
 CC APPLICANT: Pollock, Joanne  
 CC APPLICANT: Bond, Julian F.  
 CC APPLICANT: Kuo, Mei-Chang  
 CC APPLICANT: Brauer, Richard D.  
 CC APPLICANT: Yeung, Siu-mei H.  
 CC APPLICANT: Brauer, Andrew

CC APPLICANT: Griffith, Irwin J.  
 CC APPLICANT: Pollock, Joanne  
 CC APPLICANT: Bond, Julian F.  
 CC APPLICANT: Kuo, Mei-Chang  
 CC APPLICANT: Brauer, Richard D.  
 CC APPLICANT: Yeung, Siu-mei H.  
 CC APPLICANT: Brauer, Andrew  
 CC APPLICANT: Exley, Mark A.  
 CC APPLICANT: Powers, Steven F.  
 CC TITLE OF INVENTION: Allergic Proteins And Peptides From  
 CC TITLE OF INVENTION: Japanese Cedar Pollen  
 CC NUMBER OF SEQUENCES: 261  
 CC CORRESPONDENCE ADDRESS:  
 CC ADDRESSEE: Immunologic Pharmaceuticals Corporation, Inc.  
 CC STREET: 610 Lincoln St  
 CC CITY: Waltham  
 CC STATE: MA  
 CC COUNTRY: USA  
 CC ZIP: 02154  
 CC COMPUTER READABLE FORM:  
 CC MEDIUM TYPE: Floppy disk  
 CC COMPUTER: IBM PC compatible  
 CC OPERATING SYSTEM: PC-DOS/MS-DOS  
 CC SOFTWARE: Patent Release #1.0, Version #1.25  
 CC CURRENT APPLICATION DATA:  
 CC APPLICATION NUMBER: US/08/467,023  
 CC FILING DATE: June 6, 1995  
 CC CLASSIFICATION: 424  
 CC PRIORITY APPLICATION DATA:  
 CC APPLICATION NUMBER: 05/350,225  
 CC FILING DATE: December 6, 1994  
 CC ATTORNEY/AGENT INFORMATION:  
 CC NAME: Jane E. Remillard  
 CC REGISTRATION NUMBER: 38,872  
 CC REFERENCE/DOCKET NUMBER: 025.5 US02 (01-0287012)  
 CC TELECOMMUNICATION INFORMATION:  
 CC TELEPHONE: (617) 227-1400  
 CC TELEFAX: (617) 227-5941  
 CC INFORMATION FOR SEQ ID NO: 2:  
 CC SEQUENCE CHARACTERISTICS:  
 CC LENGTH: 374 amino acids  
 CC TYPE: amino acid  
 CC TOPOLOGY: linear  
 CC MOLECULE TYPE: Protein  
 CC SEQUENCE 374 AA: 4072 MW: 757.70N  
 CC Query Match 28.8% Score 2961 DB 10 Length 374  
 CC Best Local Similarity 73.6% Pct. Id. 34.0e-16  
 CC Matches 39: Conservative 7: Indels 2: Gaps 2:  
 CC 87 RPLWTFSSNNIKKPKMYIAGYKTFDGRGAV-YIGNSGPCVF\*KRVSNI 158  
 CC 73 RPLWTFSSNNIKKPKMYIAGYKTFDGRRAEVSYSVHNGA-KFIRRVGGII 124  
 CC RESULT 13  
 CC ID US-08-468-940-2 STANDARD: PRT: 374 AA.  
 CC XX xxxxxx  
 CC AC  
 CC DT  
 CC DE  
 CC Sequence 2, Application: US/08468940  
 CC Sequence 2, Application: US/08468940  
 CC GENERAL INFORMATION:  
 CC APPLICANT: Griffith, Irwin J.  
 CC APPLICANT: Pollock, Joanne  
 CC APPLICANT: Bond, Julian F.  
 CC APPLICANT: Kuo, Mei-Chang  
 CC APPLICANT: Brauer, Richard D.  
 CC APPLICANT: Yeung, Siu-mei H.  
 CC APPLICANT: Brauer, Andrew



```

00 APPLICATION NUMBER: US/08/350.225
01 FILING DATE: December 6, 1994
02 CLASSIFICATION: 424
03 PRIOR APPLICATION DATA:
04 APPLICATION NUMBER: 08/225.248
05 FILING DATE: April 8, 1994
06 APPLICATION NUMBER: 07/518.932
07 FILING DATE: September 1, 1992
08 APPLICATION NUMBER: PCI/US94/50129
09 FILING DATE: January 15, 1993
10 ATTORNEY/AGENT INFORMATION:
11 NAME: Carlene A. Vassilone
12 REGISTRATION NUMBER: 35,729
13 REFERENCE/DRAWER NUMBER: 025.6 US (IMI-C28CP2)
14 TELECOMMUNICATION INFORMATION:
15 TELEPHONE: (617) 466-6000
16 TELEFAX: (617) 466-6040
17 INFORMATION FOR SEQ ID NO: 2:
18 SEQUENCE CHARACTERISTICS:
19 LENGTH: 374 amino acids
20 TYPE: amino acid
21 TOPOLOGY: linear
22 FEATURE TYPE: PROTEIN
23 SEQUENCE 374 AA: 40720 MW: 75477 GN:
24 Query Match: 29.8% Score 296: DB #1 Length 374:
25 Best Local Similarity 73.6% Pred No 3,148-18:
26 Matches 39: Conserved: 5: Mismatches 7: Indels 2:
27
28 87 KPAWITTSQNNNKKKPMYTAGYKTFDGRGAQVYVGNSSPGVFTKRVNV: 138
29 I IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII
30 74 KPAWITTSQNNNKKKPMYTAGYKTFDGRGAQVYVGNVCAKPTREVGGY: 124
31 I IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII

```

Search completed: Mon Jun 19 16:21:59 2006  
Run time : 44 secs.

\*\*\*\*\*  
WATERBURY (TM)  
\*\*\*\*\*  
Release 3.1A John F. Collins, Bioconjugating Research Unit,  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd  
Msearch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Mon Jun 19 15:58:04 2000; Maspar time 11.79 Seconds  
Tabular output not generated.

Title: XG-09-142-524A-3  
Description: (1-134) from USC9142524A.pep  
Perfect Score: 1026  
Sequence: 1 MKVTVAFCGPNRRVFIKR.....KPIRAVDGIIAAYONPASWK 134  
Scoring table: EAM 150  
Gap 11  
Searched: 142080 seqs, 47172405 residues  
Post-processing: Minimum Match 0%  
Listing first 45 summaries  
Database: pirf2  
1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 43.199; Variance 89.328; scale 0.484  
Prod. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.  
SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description  | Prod. No. |
|------------|-------|-------------|--------|-------|--|-----------|
| 1          | 236   | 28.5        | 374    | 2     | JC2124 major allergen Cry j                        | 4.85e-33  |
| 2          | 236   | 28.5        | 374    | 2     | JC2123 major allergen Cry j                        | 4.85e-33  |
| 3          | 172   | 16.8        | 514    | 2     | JC2498 second major allergen Cry j II protein - Ja | 1.24e-11  |
| 4          | 172   | 16.8        | 514    | 2     | S49730 pectate lyase (EC 4.2.6.1) - Ja             | 1.24e-11  |
| 5          | 155   | 15.1        | 542    | 2     | T06728 pectate lyase (EC 4.2.6.1) - O9             | 6.01e-09  |
| 6          | 152   | 14.8        | 397    | 2     | S26211 pectate lyase (EC 4.2.6.1) - O8             | 1.75e-08  |
| 7          | 143   | 13.9        | 434    | 2     | S29612 pectate lyase (EC 4.2.6.1) - O7             | 4.13e-07  |
| 8          | 141   | 13.7        | 396    | 2     | A39095 allergen Amb a 1.1 pr                       | 8.25e-07  |
| 9          | 141   | 13.7        | 397    | 2     | C39099 allergen Amb a 1.3 pr                       | 8.25e-07  |
| 10         | 141   | 13.7        | 397    | 2     | C53240 allergen Amb a 1.3 pr                       | 8.25e-07  |
| 11         | 141   | 13.7        | 455    | 2     | T00856 pectate lyase (EC 4.2.6.1) - O7             | 8.25e-07  |
| 12         | 139   | 13.5        | 392    | 2     | D53240 allergen Amb a 1.4 pr                       | 1.64e-06  |
| 13         | 138   | 13.3        | 398    | 2     | T07058 pectate lyase (EC 4.2.6.1) - O6             | 2.31e-06  |
| 14         | 137   | 13.4        | 398    | 2     | B39099 allergen Amb a 1.2 pr                       | 3.26e-06  |
| 15         | 137   | 13.4        | 474    | 2     | S12209 pectate lyase (EC 4.2.6.1) - O6             | 3.26e-06  |
| 16         | 135   | 13.2        | 398    | 2     | B53240 allergen Amb a 1.2 pr                       | 6.43e-06  |
| 17         | 133   | 13.0        | 404    | 2     | T05556 pectate lyase (EC 4.2.6.1) - O5             | 2.27e-05  |
| 18         | 131   | 12.8        | 397    | 2     | A04659 allergen Amb alpha II                       | 2.48e-05  |
| 19         | 131   | 12.8        | 397    | 2     | S53240 allergen Amb a II pre                       | 2.48e-05  |
| 20         | 131   | 12.8        | 438    | 2     | S43335 pectate lyase (EC 4.2.6.1) - O5             | 2.48e-05  |
| 21         | 130   | 12.7        | 450    | 2     | T09524 probable pectate lyase                      | 3.46e-05  |
| 22         | 125   | 12.2        | 418    | 2     | T07701 pectate lyase (EC 4.2.6.1) - O4             | 1.81e-04  |
| 23         | 124   | 12.1        | 274    | 2     | T03932 pectate lyase (EC 4.2.6.1) - O4             | 2.52e-04  |

|    |     |      |      |   |  |
|----|-----|------|------|---|--|
| 24 | 124 | 12.1 | 374  | 2 | T05240 pectate lyase (EC 4.2.6.1) - O4 |
| 25 | 124 | 12.1 | 449  | 2 | S27088 pectate lyase (EC 4.2.6.1) - O4 |
| 26 | 110 | 10.7 | 4450 | 2 | JX0340 gramicidin S synthase           |
| 27 | 105 | 10.2 | 110  | 2 | D48562 hypothetical protein            |
| 28 | 100 | 9.7  | 238  | 2 | S23114 hypothetical protein            |
| 29 | 100 | 9.7  | 333  | 2 | T01456 hypothetical protein            |
| 30 | 98  | 9.6  | 1084 | 2 | S23319 hypothetical protein            |
| 31 | 99  | 9.6  | 2092 | 2 | S33236 genome polyprotein              |
| 32 | 99  | 9.6  | 2149 | 2 | S13376 genome polyprotein              |
| 33 | 96  | 9.4  | 111  | 2 | F3352 hypothetical protein             |
| 34 | 96  | 9.4  | 190  | 2 | E7288 hypothetical protein             |
| 35 | 96  | 9.4  | 1291 | 1 | S55455 retrovirus-related po           |
| 36 | 95  | 9.3  | 1070 | 2 | S27555 hypothetical protein            |
| 37 | 95  | 9.3  | 3225 | 2 | S27852 probable cell-surface           |
| 38 | 94  | 9.2  | 554  | 2 | S45303 hypothetical protein            |
| 39 | 93  | 9.1  | 539  | 2 | F70737 chaperonin GroEL - M            |
| 40 | 93  | 9.1  | 875  | 1 | A57080 cell surface antigen            |
| 41 | 92  | 9.0  | 215  | 2 | S64590 a2-mating type protei           |
| 42 | 91  | 8.9  | 131  | 2 | I49566 hypothetical protein            |
| 43 | 91  | 8.9  | 216  | 2 | T03347 hypothetical protein            |
| 44 | 91  | 8.9  | 426  | 2 | T07810 silicosis-specific gly          |
| 45 | 91  | 8.9  | 596  | 2 | T07421 phosphoinositide-spec           |

ALIGNMENTS

RESULT 1  
ENTRY JC2124 \*type complete  
TITLE major allergen Cry j I precursor (clone pCC1-15) - Japanese cedar  
ORGANISM Cedar  
DATE 14-Jul-1994 \*sequence\_revision 14-Jul-1994 \*text\_change 26-Aug-1999  
ACCESSION JC2124  
REFERENCE JC2123  
#authors Sone, T.; Komiyama, N.; Shimizu, K.; Kusakabe, T.; Marikubo, K.; Kino, K.  
#journal Biochem. Biophys. Res. Commun. (1994) 199:19-425  
#title Cloning and sequencing of cDNA coding for Cry j I a major allergen of Japanese cedar pollen.  
#cross-references MJD:9418334  
#accession JC2124  
#molecule\_type mRNA  
#residues 1-374 \*label SON  
#cross-references GB:528545; NID:494633; PID:405072; PID:437434  
#experimental\_source pollen  
#note The authors designated this sequence as Cry j I residue 219

CLASSIFICATION \*serpin family 1 peptidase-like  
KEYWORDS glycoprotein; pollen  
FEATURE  
22-374 \*domain signal sequence \*status predicted \*label SON  
\*product major allergen Cry j I (clone pCC1-15) \*status predicted \*label MAN  
158-191-293-354 \*binding site carbohydrate (Asn) (covalent) \*status predicted  
SUMMARY \*length: 374 \*molecular-weight 40702 \*checksum 3592  
Query Match: 28.8% Score 296; DB 2: Length 374;  
Best Local Similarity 73.5%; Pred. No. 4.86e-33;  
Matches 39; Conservative 5; Mismatches 7; Indels 2; Gaps 2;

Db 87 RPLWIFSGNNKIKLMPYIAGYKTFDGRCAOV-YICNSGPCVFIKRVSNV: 139  
|||||  
QY 73 RPLWIFSGNNKIKLMPYIAGYKTFDGRRAEVSYVHVNSA-KPIRRVSGI: 124  
|||||  
RESULT 2  
ENTRY JC2123 \*type complete  
TITLE major allergen Cry j I precursor (clone pCC1-2-2) - Japanese cedar  
ORGANISM Cedar  
DATE 14-Jul-1994 \*sequence\_revision 14-Jul-1994 \*text\_change

```

26-Apr-1999
J02123: PC2065
REFERENCE
  J02123
  Sone, T.; Komiya, N.; Shimizu, K.; Kusakabe, T.; Morikubo,
  K.; Kito, K.
  Biochem Biophys Res Commun. (1994) 199:519-525
  Cloning and sequencing of cDNA coding for Cry 1 I, a major
  allergen of Japanese cedar pollen.
  cross-references MIM:342532
  accession J02123
  molecule_type mRNA
  residues 1-374 #label SGN
  cross-references GB:26544; NID:043631; PID:000085; PID:0493632
  expofrontal_source pollen
  accession PC2065
  molecule_type protein
  residues 22-53; 58-91; 219-232; 236-248; 259-307; 345-372 #label SC2
  note the authors described car: hydrate binding site for
  residue 273
CLASSIFICATION
  superfamily: pectate lyase LA159
  keywords
  allergen; protein
  domain: signal sequence
  status: predicted #label SGN
  product: major allergen Cry 1 I (clone PC2-2-2) #status
  predicted #label MA1
  binding_site carbohydrate (ASN) (covariant) #status
  predicted
SUMMARY
  length 374 #molecular-weight 40545 #checksum 2920
  Query Match 28.8% Score 296; DB 2; Length 374;
  Best Local Similarity 73.6% Pred. No. 4; 856-33;
  Matches 39; Conservative 5; Mismatches 7; Indels 2; Gaps 2;
  DB 87 RPKWIFSNMNIKKMPYIAGYKTFDGRAGV-VIOMGGPCVFIKVSNNV 138
  QY 73 RPKWIFSNMNIKKMPYIAGYKTFDGRRAEVSIVHNGA-KFIRVQNI 124
  RESULT 3
  ENTRY
  TITLE
  ORGANISM
  DATE
  12-Sep-1997
  accession J02498; PC2346; A60147
  J02498
  Komiya, N.; Sone, T.; Shimizu, K.; Morikubo, K.; Kito, K.
  Biochem Biophys Res Commun. (1994) 201:1021-1028
  cDNA cloning and expression of Cry 1 II, the second major
  allergen of Japanese cedar pollen.
  cross-references MIM:342185
  accession J02498
  molecule_type mRNA
  residues 1-514 #label KOM
  cross-references GB:129772; NID:0506857; PID:0005734; PID:0506859
  accession PC2346
  molecule_type protein
  residues 52-61 #label KO2
  accession A60147
  authors
  Sakaue, M.; Ino, Y.; Saito, T.; Tanaka, M.; Ando, S.; Ueda, M.;
  Matsuda, T.
  Allergy (1993) 45:309-312
  Identification of the second major allergen of Japanese cedar
  pollen.
  cross-references MIM:342358
  accession A60147
  molecule_type protein
  residues 55-64 #label SAK
  keywords
  glycoprotein; pollen
  domain: signal sequence
  status: predicted #label SGN
  product: second major allergen Cry 1 #status: predicted
  #label MA1

```

```

429,450,472
SUMMARY
  length 514 #molecular-weight 55683 #checksum 472
  Query Match 16.6% Score 172; DB 2; Length 514;
  Best Local Similarity 95.2% Pred. No. 1; 246-11;
  Matches 20; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
  DB 133 WKNNRWLQFAKLITFLMKR 159
  QY 52 WKNNRWLQFAKLITFLMKR 72
  RESULT 4
  ENTRY
  TITLE
  ORGANISM
  DATE
  28-May-1995
  accession S48730
  S48730
  Namba, M.; Kurase, M.; Taniuchi, K.; Hino, K.; Yamashita, Y.;
  Fukuda, S.; Usui, M.; Katamoto, M.
  PEST Lett. (1994) 33:124-128
  Molecular cloning of the second major allergen, Cry 1 II,
  from Japanese cedar pollen.
  cross-references MIM:0501077
  accession S48730
  status: preliminary
  molecule_type mRNA
  residues 1-514 #label NAM
  cross-references GB:37765; NID:057655; PID:0002210;
  PID:0007598; PID:057656
SUMMARY
  length 514 #molecular-weight 56445 #checksum 54
  Query Match 16.8% Score 172; DB 2; Length 514;
  Best Local Similarity 95.2% Pred. No. 1; 246-11;
  Matches 20; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
  DB 133 WKNNRWLQFAKLITFLMKR 159
  QY 52 WKNNRWLQFAKLITFLMKR 72
  RESULT 5
  ENTRY
  TITLE
  ORGANISM
  DATE
  28-Apr-1995
  accession J06728
  J06728
  Quetier, F.; Choisme, N.; Robert, C.; Brothier, P.; Winkler,
  P.; Catala, J.; Artiguenave, F.; Saito, M.;
  Weissenbach, J.; Salasubbat, M.; Mewes, H.W.; Mayer,
  K.F.X.; Scheller, G.
  submitted to the Protein Sequence Database, April 1995
  accession J06728
  molecule_type DNA
  residues 1-542 #label GCF
  cross-references EMBL:A049655
  expofrontal_source cDNA; Columbia; BAC clone F28PL
  GENETICS
  map_position 3
  introns
  note
  F28P10.100
  classification
  superfamily: pectate lyase LA159
  keywords
  carboxyl-lyase
  summary
  length 542 #molecular-weight 56573 #checksum 1032
  Query Match 15.1% Score 155; DB 2; Length 542;

```

```

Best Local Similarity 52.6%; Pred. No. 6,01e-09;
Matches 19; Conservative 7; Mismatches 7; Indels 0; Gaps 0;

DB 146 PLMTVFSSNMLIRKQHLINYSYKIDGRSAV 178
      |||||  |||||  |||||  |||||  |||||
QY 74 PLMTFSSNNMLKMKPMYAGYKTFDGRRAEV 106

RESULT 6
ENTRY pectate lyase (EC 4.2.2.2) - common tobacco
TITLE pectate lyase (EC 4.2.2.2) - common tobacco
CKANISM #forma_name Nicotiana glauca common tobacco
DATE 12-Feb-1993 #sequence_revision 12-Feb-1993 #text_change
29-Sep-1999
ACCESSIONS S26211 S26212 S21933 S22754
REFERENCE S26211
#authors Rogers, H.J.; Harvey, A.; Lonsdale, D.M.
#journal Plant Mol. Biol. (1992) 20:493-502
#title Isolation and characterization of a tobacco gene with
#title homology to pectate lyase which is specifically expressed
#title during microsporogenesis.
#cross-references MIM:31043035
#accession S26211
#molecule_type DNA
#residues 1-397 #label MOG
#cross-references EMBL:X67158; NID:g19947; PIDN:CAA47630.1; PID:g19908
#accession S26212
#molecule_type mRNA
#residues 119-155; C 157-188; G 191-199; D 200; R 203-248;
#cross-references EMBL:X67159
#note Translation of the nucleotide sequence is not complete
REFERENCE S21933
#authors Lonsdale, D.M.
#submission Submitted to the EMBL Data Library, July 1991
#accession S21933
#molecule_type DNA
#residues 1-397 #label LON
#cross-references EMBL:X61102; NID:g19981; PIDN:CAA43414.1; PID:g19982
GENETICS 193/1: 293/2
#initiator superfamily pectate lyase LAT59
CLASSIFICATION carbon-oxygen lyase
KEYWORDS #length 397 #molecular-weight 44351 #checksum 1480
SUMMARY
Query Match 11.8%; Score 152; DB 2; Length 397;
Best Local Similarity 36.7%; Pred. No. 1.75e-08;
Matches 22; Conservative 15; Mismatches 22; Indels 1; Gaps 1;

DB 105 YGVICRPLWIFGNMKIKSLKELIVSNKTIIDRGFNVIHUNAGTKTQASNTIIS 163
      |||||  |||||  |||||  |||||  |||||
QY 67 FTLMGRPLWIFSSNNMLKMKPMYIAYKTFDGRRAEVSVHVNAGKTRVVGIIAA 126

RESULT 7
ENTRY pectate lyase (EC 4.2.2.2) - trumpet lily
TITLE pectate lyase (EC 4.2.2.2) - trumpet lily
CKANISM #forma_name Lilium longiflorum #common_name trumpet lily
DATE 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change
29-Sep-1999
ACCESSIONS S29612
REFERENCE S29611
#authors Kim, S.; Finkel, D.J.; An, G.
#submission Submitted to the EMBL Data Library, October 1992
#description Abundancy patterns of lily pollen cDNAs: characterization of
#description three pollen-preferential cDNA clones.
#accession S29612
#status preliminary
#molecule_type mRNA
#residues 1-434 #label KIM
#cross-references EMBL:Z17328; NID:g19450; PIDN:CAA78976.1; PID:g19451
#experimental_source cv. Nellie White, mature flower
CLASSIFICATION #superfamily pectate lyase LAT59

```

```

KEYWORDS carbon-oxygen lyase
SUMMARY #length 434 #molecular-weight 48457 #checksum 6472

Query Match 13.9%; Score 143; DB 2; Length 434;
Best Local Similarity 39.5%; Pred. No. 4.13e-07;
Matches 25; Conservative 14; Mismatches 22; Indels 4; Gaps 4;

DB 140 FGTLRWGVICRPLWIFGNMKIKSLKELIVSNKTIIDRGFNVIHUNAGKTRVVGIIAA 163
      |||||  |||||  |||||  |||||  |||||
QY 61 FAKLTGFTLMGRPLWIFSSNNMLKMKPMYIAYKTFDGRRAEVSVHVNAGKTRVVGIIAA 124

DB 197 VHNVI 201
      |
QY 120 VDSII 124

RESULT 8
ENTRY #type complete
TITLE allergen Amb a 1 precursor - common ragweed
ORGANISM #forma_name Artemisia artemisiifolia #common_name ragweed
DATE 27-Nov-1991 #sequence_revision 03-Apr-1992 #text_change
29-Sep-1999
ACCESSIONS A39099 A60895 A53240
REFERENCE A39099
#authors Rainey, T.; Griffith, L.J.; Kuo, M.; Bond, J.F.; Roberts,
#journal J. Biol. Chem. (1991) 266:1229-1236
#title Cloning of Amb a 1 (allergen E), the major allergen family of
#title short ragweed pollen.
#cross-references MIM:3103235
#accession A39099
#molecule_type mRNA
#residues 1-396 #label RAF
#cross-references GB:M53116
REFERENCE A60895
#authors Smith, J.J.; Olson, J.K.; Klapper, E.G.
#journal Mol. Immunol. (1988) 25:155-165
#title Monoclonal antibodies to denatured ragweed pollen allergen
#title Amb a 1: characterization, specificity for the denatured
#title allergen, and utilization for the isolation of immunogenic
#title peptides of Amb a 1.
#cross-references MIM:3103306 #label SM1
#accession A60895
#molecule_type protein
#residues 256-273; 252-303; 305-306 #label SM1
REFERENCE A53240
#authors Griffith, L.J.; Clark, J.; Klapper, E.G.; Roberts, P.L.;
#journal J. Allergy Clin. Immunol. (1992) 89:69-84
#title Sequence polymorphism of Amb a 1 and Amb a 2 in the major
#title allergens in Artemisia artemisiifolia (Sant. Rudbeck)
#cross-references MIM:3103470
#accession A53240
#status preliminary
#molecule_type mRNA
#residues 1-396 #label GR1
#cross-references GB:M5358; NID:g166434; PIDN:AA3265.1; PID:g166435
CLASSIFICATION #superfamily pectate lyase LAT59
KEYWORDS glycoprotein; pollen
FEATURE
36 #binding-site carbohydrate (Asn) (covalent) #status
predicted
SUMMARY #length 396 #molecular-weight 42709 #checksum 5506

Query Match 13.7%; Score 141; DB 2; Length 396;
Best Local Similarity 43.2%; Pred. No. 8.25e-07;
Matches 19; Conservative 8; Mismatches 17; Indels 0; Gaps 0;

DB 107 FGAAQNRPLWIFERDMVIRLQKGVVNSDKTIDGRGAKVEIIN 150
      |||||  |||||  |||||  |||||  |||||
QY 67 FTLMGRPLWIFSSNNMLKMKPMYIAYKTFDGRRAEVSVH 110

```

```

RESULT 1
ENTRY
  accession 039099 #type complete
  title allergen Amb a 1.3 - common ragweed
  organism Arabidopsis thaliana #common_name common
  date 27-Nov-1994 #sequence_revision 03-Apr-1992 #text_change
  26-Sep-1992
ACCESSIONS
  reference A39099
  authors Rainard, J., Griffith, L.J., Kuo, M., Bond, J.E., Rogers, B.L., Klapper, D.G.
  journal J. Biol. Chem. (1991) 266:1229-1236
  title Cloning of Amb a 1 (allergen E), the major allergen family of
  start ragweed pollen.
  cross-references M01D:Q16325
  accession 039099
  status preliminary
  molecule_type mRNA
  residues 1-397 #label EAF
  cross-references G5:K52961; NID:Q16444; P1DN:AAA32658.1; P1D:3:56441
  classification #superfamily pectate lyase LA159
  keywords pollen
  summary #length 397 #molecular-weight 42962 #checksum 9775
  query match 13.7% Score 141; DB 2; Length 397;
  best local similarity 37.9% Pred. No. 8,256-07;
  matches 22; Conservative 12; Mismatches 22; Indels 2; Gaps 2;

DB 108 FGAQNRPLWIFKQKVIINLNGELVYVNSDKTIIGRGVKVE-I-INGGLIMNVKNI 163
||||| 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100
67 FTLMGRPLWIFSGNNKIKKPMYIAGYKFDGRAEVSVYVHVGAKFIRVCGII 124

RESULT 12
ENTRY
  accession 053240 #type complete
  title allergen Amb a 1.3 precursor - common ragweed
  organism Arabidopsis thaliana #common_name common
  date 19-May-1994 #sequence_revision 19-May-1994 #text_change
  18-Sep-1994
ACCESSIONS
  reference A53240
  authors Griffith, L.J., Pollock, J., Klapper, D.G., Rogers, B.L., Nault, A.K.
  journal Int. Arch. Allergy Appl. Immunol. (1991) 96:296-304
  title Sequence polymorphism of Amb a 1 and Amb a 11, the major
  allergens in Arabidopsis thaliana (short ragweed).
  cross-references M01D:Q16452
  accession 053240
  status preliminary
  molecule_type mRNA
  residues 1-397 #label GRT
  cross-references G5:M0562; NID:Q16444; P1DN:AAA32658.1; P1D:3:56445
  classification #superfamily pectate lyase LA159
  keywords pollen
  summary #length 397 #molecular-weight 42963 #checksum 9775
  query match 13.7% Score 141; DB 2; Length 397;
  best local similarity 37.9% Pred. No. 8,256-07;
  matches 22; Conservative 12; Mismatches 22; Indels 2; Gaps 2;

DB 108 FGAQNRPLWIFKQKVIINLNGELVYVNSDKTIIGRGVKVE-I-INGGLIMNVKNI 163
||||| 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100
67 FTLMGRPLWIFSGNNKIKKPMYIAGYKFDGRAEVSVYVHVGAKFIRVCGII 124

RESULT 13
ENTRY
  accession T07058 #type complete
  title pectate lyase (EC 4.2.2.2) LA156 - tomato
  organism Arabidopsis thaliana #common_name tomato
  date 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change

```

```

ACCESSIONS
  reference T0856
  authors Roundsley, S.D., Lin, X., Ketchum, K.A., Grosly, M.,
  Stradon, R.C., Sykes, S.M., Kaul, S., Mason, I.M.,
  Kertlavade, A.R., Adams, M.D., Schervillo, J.P., Vetter,
  J.C.
  submission submitted to the EMBL Data Library, March 1998
  description Arabidopsis thaliana chromosome 11: BAC T20F6.14
  sequence.
  accession T0856
  status translated from G5/EMBL/DBA
  molecule_type DNA
  residues 1-455 #label RQI
  cross-references EMBL:AC02621; NID:Q247056; P1DN:AAA32658.1
  experimental_source coll. Columbia
  map_position 2
  introns 65/2; 295/3; 376/3
  note T20F6.14
  classification #superfamily pectate lyase LA159
  keywords carbon:oxygen lyase
  summary #length 455 #molecular-weight 51257 #checksum 911
  query match 13.7% Score 141; DB 2; Length 455;
  best local similarity 45.1% Pred. No. 8,256-07;
  matches 23; Conservative 9; Mismatches 15; Indels 5; Gaps 4;

DB 156 RPLWIFARSWIKGCELIINDKTIIGRGVKVE-I-INGGLIMNVKNI 215
||||| 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100
73 RPLWIFSGNNKIKKPMYIAGYKFDGRAEVSVYVHVGAKFIRVCGII 124

RESULT 12
ENTRY
  accession 053240 #type complete
  title allergen Amb a 1.3 precursor - common ragweed
  organism Arabidopsis thaliana #common_name common
  date 19-May-1994 #sequence_revision 19-May-1994 #text_change
  26-Sep-1994
ACCESSIONS
  reference A53240
  authors Griffith, L.J., Pollock, J., Klapper, D.G., Rogers, B.L., Nault, A.K.
  journal Int. Arch. Allergy Appl. Immunol. (1991) 96:296-304
  title Sequence polymorphism of Amb a 1 and Amb a 11, the major
  allergens in Arabidopsis thaliana (short ragweed).
  cross-references M01D:Q16452
  accession 053240
  status preliminary
  molecule_type mRNA
  residues 1-397 #label GRT
  cross-references G5:M0562; NID:Q16444; P1DN:AAA32658.1; P1D:3:56445
  classification #superfamily pectate lyase LA159
  keywords pollen
  summary #length 397 #molecular-weight 42842 #checksum 9775
  query match 13.5% Score 139; DB 2; Length 397;
  best local similarity 36.4% Pred. No. 1,646-56;
  matches 20; Conservative 11; Mismatches 23; Indels 1; Gaps 1;

DB 108 FGAQNRPLWIFKQKVIINLNGELVYVNSDKTIIGRGVKVE-I-INGGLIMNVKNI 162
||||| 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100
67 FTLMGRPLWIFSGNNKIKKPMYIAGYKFDGRAEVSVYVHVGAKFIRVCGII 124

RESULT 13
ENTRY
  accession T07058 #type complete
  title pectate lyase (EC 4.2.2.2) LA156 - tomato
  organism Arabidopsis thaliana #common_name tomato
  date 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change

```









AT THEIR NON-REDUCING ENDS.  
 -1- DEVELOPMENTAL STAGE: MAXIMALLY EXPRESSED LATE IN POLLEN DEVELOPMENT.  
 -1- SIMILARITY: BELONGS TO THE POLYSACCHARIDE LYASE FAMILY 1.  
 -----  
 This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----  
 DR EMBL: X67158; CAA47630.1; .  
 DR EMBL: X67159; CAA47631.1; .  
 DR EMBL: X67160; CAA47632.1; .  
 DR PIR: S26211; S26211.  
 DR PFAM: PF00544; pec\_lyase.1.  
 DR PRINTS: PR00807; ANBALLERG.  
 DR SIGNAL: 1 25 POTENTIAL.  
 FT CHAIN 26 357 PECTATE LYASE.  
 FT ACT\_SITE 272 272 POTENTIAL.  
 FT CARBOHYD 134 134 POTENTIAL.  
 FT CARBOHYD 227 227 POTENTIAL.  
 FT CONFLICT 156 156 S -> C (IN MRNA).  
 FT CONFLICT 189 190 GS -> SG (IN MRNA).  
 FT CONFLICT 200 200 S -> D (IN MRNA).  
 FT CONFLICT 202 202 H -> R (IN MRNA).  
 FT CONFLICT 249 249 H -> N (IN MRNA).  
 SQ SEQUENCE 367 AA: 4435; MW: 5048255DA7643F CRC64;  
 -----  
 Query Match 14.8%; Score 152; DB 1; Length 397;  
 Best Local Similarity 36.7%; Pred. No. 3.64e-10;  
 Matches 22; Conservative 15; Mismatches 22; Indels 1; Gaps 1;  
 -----  
 DR 125 YVYVCKENLIIIFKNNKIKLSRELIVTSKTIIDGRFNVHIONGAGIK-IQASNIILIS 153  
 CY 67 ITLKGRRPLWIFKNNKIKLMKMYIAGYKTFDGRRAEVSYYVHVNCAGKFRVVDGIIAA 126  
 -----  
 RESULT 4  
 ID PELLILLO STANDARD: 981; 434 AA.  
 AC 146978;  
 DT 01-FEB-1995 (Rel. 31; Created);  
 DT 01-FEB-1995 (Rel. 31; Last sequence update);  
 DT 15-DEC-1998 (Rel. 37; Last annotation update);  
 DE PECTATE LYASE PRECURSOR (EC 4.2.2.2).  
 LS Lilium longiflorum (Trumpet Lily).  
 CC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta;  
 CC Euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;  
 CC core eudicots; Asteridae; euasterids I; Asterales; Asteroideae;  
 CC Liliaceae; Liliaceae.  
 RN [1]  
 RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.  
 RC TISSUE-POLLEN;  
 RX MEDLINE: 91093235.  
 RA Rafnar T., Griffith I.J., Kuo X.-C., Bond J.F., Rogers B.,  
 RA Klapper D.G.;  
 RT "Cloning of Amb 4 (antigen E), the major allergen family of Stort  
 RT ragweed pollen.";  
 RL J. Biol. Chem. 266:1229-1236(1991).  
 RN [2]  
 RP SEQUENCE FROM N.A. AND VARIANTS  
 RC TISSUE-POLLEN;  
 RX MEDLINE: 92234570.  
 RA Griffith I.J., Pollock J., Klapper D.G., Rogers B., Nall A.K.;  
 RT "Sequence polymorphism of Amb 4 and Amb 5 in ragweed pollen and  
 RT to Ambrosia alternifolia (Ragwort) pollen.";  
 RL Int. Arch. Allergy Appl. Immunol. 102:44-49(1991).  
 CC -1- SUBUNIT: MONOMER.  
 CC -1- TISSUE SPECIFICITY: POLLEN AND FLOWERS  
 CC -1- DISEASE: ONE OF THE MAJOR ALLERGENS OF THE RAGWEED POLLEN.  
 CC -1- SIMILARITY: BELONGS TO THE POLYSACCHARIDE LYASE FAMILY 1.  
 CC AMB 4/AMB 5 II/CRY 3 I SUBFAMILY.  
 -----  
 This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----  
 DR EMBL: M63116; .; NOT-ANNOTATED\_CDS.  
 DR EMBL: M80558; AAA32655.1; .  
 DR PIR: A39099; A39099.  
 DR PIR: A53240; A53240.  
 DR PFAM: PF00544; pec\_lyase.1.  
 DR PRINTS: PR00807; ANBALLERG.  
 DR Antigen: Allergen; Signal: Multigene family; Polymorphism.  
 KW

DR PFAM: PF00544; pec\_lyase.  
 DR PRINTS: PR00807; ANBALLERG  
 DR Lyase: SIGNAL: 1 22 POTENTIAL.  
 FT CHAIN 23 434 PECTATE LYASE.  
 FT ACT\_SITE 312 312 POTENTIAL.  
 FT CARBOHYD 68 68 POTENTIAL.  
 FT CARBOHYD 97 97 POTENTIAL.  
 SQ SEQUENCE 434 AA: 48457 MW: 51936040H80C64 CRC64;  
 -----  
 Query Match 13.9%; Score 143; DB 1; Length 444;  
 Best Local Similarity 38.5%; Pred. No. 1.32e-08;  
 Matches 25; Conservative 14; Mismatches 22; Indels 4; Gaps 4;  
 -----  
 DR 140 FGLUR-MGVIOQRPLWIFKNNKIKLMKMYIAGYKTFDGRRAEVSYYVHVNCAGKFRVVDGIIAA 196  
 CY 61 FAKLTGFLMGRPLWIFSQ :KIKMRYIAGYKTFDGRRAEVSYYVHVNCAGKFRVVDGIIAA 119  
 -----  
 DR 197 VHVVI 201  
 CY 120 VDGII 124  
 -----  
 RESULT 5  
 ID MP11-AMBAR STANDARD: 91; 396 AA.  
 AC P27759;  
 DT 01-AUG-1992 (Rel. 23; Created);  
 DT 01-AUG-1992 (Rel. 23; Last sequence update);  
 DT 15-DEC-1998 (Rel. 37; Last annotation update);  
 DE POLLEN ALLERGEN AMB A 1.1 PRECURSOR (ANTIGEN E) (ANTIGEN AMB A 1).  
 OS Ambrosia artemisiifolia (Stor asweed).  
 CC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta;  
 CC Euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;  
 CC core eudicots; Asteridae; euasterids I; Asterales; Asteroideae;  
 CC Asteroideae; Helianthaceae; Ambrosia.  
 RN [1]  
 RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.  
 RC TISSUE-POLLEN;  
 RX MEDLINE: 91093235.  
 RA Rafnar T., Griffith I.J., Kuo X.-C., Bond J.F., Rogers B.,  
 RA Klapper D.G.;  
 RT "Cloning of Amb 4 (antigen E), the major allergen family of Stort  
 RT ragweed pollen.";  
 RL J. Biol. Chem. 266:1229-1236(1991).  
 RN [2]  
 RP SEQUENCE FROM N.A. AND VARIANTS  
 RC TISSUE-POLLEN;  
 RX MEDLINE: 92234570.  
 RA Griffith I.J., Pollock J., Klapper D.G., Rogers B., Nall A.K.;  
 RT "Sequence polymorphism of Amb 4 and Amb 5 in ragweed pollen and  
 RT to Ambrosia artemisiifolia (Ragwort) pollen.";  
 RL Int. Arch. Allergy Appl. Immunol. 102:44-49(1991).  
 CC -1- SUBUNIT: MONOMER.  
 CC -1- TISSUE SPECIFICITY: POLLEN AND FLOWERS  
 CC -1- DISEASE: ONE OF THE MAJOR ALLERGENS OF THE RAGWEED POLLEN.  
 CC -1- SIMILARITY: BELONGS TO THE POLYSACCHARIDE LYASE FAMILY 1.  
 CC AMB 4/AMB 5 II/CRY 3 I SUBFAMILY.  
 -----  
 This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----  
 DR EMBL: M63116; .; NOT-ANNOTATED\_CDS.  
 DR EMBL: M80558; AAA32655.1; .  
 DR PIR: A39099; A39099.  
 DR PIR: A53240; A53240.  
 DR PFAM: PF00544; pec\_lyase.1.  
 DR PRINTS: PR00807; ANBALLERG.  
 DR Antigen: Allergen; Signal: Multigene family; Polymorphism.  
 KW

FT SIGNAL 1 25 POTENTIAL.  
 FT CHAIN 26 396 POLLEN ALLERGEN AMB A 1.1.  
 FT MOD.RES 226 226 BLOCKED.  
 FT VARIANT 92 92 E -> D.  
 SQ SEQUENCE 396 AA: 42703 MW: 0CE75DE7B288841D CRC64:

Query Match 13.7% Score 141: DB 1: Length 396:  
 Best Local Similarity 43.2% Pred. No. 2.89e-08:  
 Matches 19: Conservative 8: Mismatches 17: Indels 0: Gaps 0:

DB 107 FGAAGNRPLWIFKNDVIRLQKVMVNSUKITIGRGAKVEIIN 150  
 QY 67 FTLMGRRLPWIFSGNNKIKMKMYIAIKYKIDGRRAEVSVMVNSAKFIKRAVLCII 149

RESULT 1  
 ID MP14\_AMBAR STANDARD: PRI: 397 AA.

AC P2751:  
 DT 01-AUG-1992 (Rel. 24, Created)  
 DT 01-DEC-1992 (Rel. 25, Last sequence update)  
 DT 15-DEC-1996 (Rel. 37, Last annotation update)  
 DE POLLEN ALLERGEN AMB A 1.3 PRECURSOR (ANTIGEN E) (ANTIGEN AMB A 1.3)  
 OS Ambrosia artemisiifolia (Short ragweed)  
 OS Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta:  
 OC euphyllophytes: Spermatophyta: Magnoliophyta: eudicotyledons:  
 OC core eudicots: Asteridae: euasterids II: Asterales: Asteraceae:  
 OC Asteroideae: Heliantheae: Ambrosia.

RN (1)  
 RP SEQUENCE FROM N.A. AND VARIANTS.  
 RC TISSUE: POLLEN.  
 RX MEDLINE: 91093245.  
 RA Raftar T., Griffith I.J., Kio M.C., Bond J.F., Rogers B.L.,  
 RA Klapper D.G.,  
 RT "Sequences of Amb a 1 (antigen E), the major allergen family of short  
 ragweed pollen." J. Biol. Chem 264:1229-1236(1991).

RN (2)  
 RP SEQUENCE FROM N.A. AND VARIANTS.  
 RC TISSUE: POLLEN.  
 RX MEDLINE: 9223457.  
 RA Griffith I.J., Pollock J., Klapper D.G., Rogers B.L., Nault A.K.,  
 RT "Sequence polymorphism of Amb a 1 and Amb a 11: the major allergens  
 in Ambrosia artemisiifolia (short ragweed)." J. All. Arch. Allergy Appl. Immunol. 9:1295-1304(1994).

CC -1- SUBUNIT: MONOMER.  
 CC -1- TISSUE SPECIFICITY: POLLEN AND FLOWERS.  
 CC -1- DISEASE: ONE OF THE MAJOR ALLERGENS OF THE RAGWEED POLLEN.  
 CC -1- SIMILARITY: BELONGS TO THE POLYSACCHARIDE LYASE FAMILY 1:  
 AMB A 1/AMB A 11/CRY J 1 SUBFAMILY

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch))  
 DR EMBL: M62956:1: AAA32659.1:1  
 DR EMBL: M62956:1: AAA32659.1:1: ALT\_SEQ.  
 DR PIR: C19099: C19099.  
 CR PFAM: PF05544: pec\_lyase: 1.  
 CR PRINTS: PR0807: AMBALLERGEN.  
 KW Allergen: Allergen: Signal: M. 1000 family: Polymorphism.  
 FT SIGNAL 1 25 POTENTIAL.  
 FT CHAIN 26 397 POLLEN ALLERGEN AMB A 1.3.  
 FT MOD.RES 226 226 BLOCKED.  
 FT VARIANT 48 48 L -> Y.  
 SQ SEQUENCE 397 AA: 42928 MW: C8DB41257590DD0A CRC64:

Query Match 13.7% Score 141: DB 1: Length 397:  
 Best Local Similarity 37.9% Pred. No. 2.89e-08:  
 Matches 22: Conservative 12: Mismatches 28: Indels 2: Gaps 2:

DB 108 FGAAGNRPLWIFKNDVIRLQKVMVNSUKITIGRGAKVEIIN 150  
 QY 67 FTLMGRRLPWIFSGNNKIKMKMYIAIKYKIDGRRAEVSVMVNSAKFIKRAVLCII 149

RESULT 7  
 ID MP14\_AMBAR STANDARD: PRI: 392 AA.

AC P29744:  
 DT 01-DEC-1992 (Rel. 24, Created)  
 DT 01-DEC-1992 (Rel. 24, Last sequence update)  
 DT 15-DEC-1996 (Rel. 37, Last annotation update)  
 DE POLLEN ALLERGEN AMB A 1.4 PRECURSOR (ANTIGEN E) (ANTIGEN AMB A 1.3)  
 OS Ambrosia artemisiifolia (Short ragweed)  
 OS Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta:  
 OC euphyllophytes: Spermatophyta: Magnoliophyta: eudicotyledons:  
 OC core eudicots: Asteridae: euasterids II: Asterales: Asteraceae:  
 OC Asteroideae: Heliantheae: Ambrosia.

RN (1)  
 RP SEQUENCE FROM N.A. AND VARIANTS.  
 RC TISSUE: POLLEN.  
 RX MEDLINE: 92234570.  
 RA Griffith I.J., Pollock J., Klapper D.G., Rogers B.L., Nault A.K.,  
 RT "Sequence polymorphism of Amb a 1 and Amb a 11: the major allergens  
 in Ambrosia artemisiifolia (short ragweed)." J. All. Arch. Allergy Appl. Immunol. 9:1295-1304(1994).

CC -1- SUBUNIT: MONOMER.  
 CC -1- TISSUE SPECIFICITY: POLLEN AND FLOWERS  
 CC -1- DISEASE: ONE OF THE MAJOR ALLERGENS OF THE RAGWEED POLLEN  
 CC -1- SIMILARITY: BELONGS TO THE POLYSACCHARIDE LYASE FAMILY 1:  
 AMB A 1/AMB A 11/CRY J 1 SUBFAMILY

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch))  
 DR EMBL: M80562:1: AAA32670.1:1  
 DR PIR: D53240:1: D53240.  
 DR PFAM: PF05544: pec\_lyase: 1.  
 DR PRINTS: PR0807: AMBALLERGEN.  
 KW Allergen: Allergen: Signal: M. 1000 family: Polymorphism.  
 FT SIGNAL 1 25 POTENTIAL.  
 FT CHAIN 26 392 POLLEN ALLERGEN AMB A 1.4.  
 FT MOD.RES 226 226 BLOCKED.  
 FT VARIANT 182 188 SHGPPV -> GNDLPPA.  
 SQ SEQUENCE 392 AA: 42342 MW: 394230021F0505A5 CRC64:

Query Match 13.5% Score 139: DB 1: Length 392:  
 Best Local Similarity 15.4% Pred. No. 1.40e-07:  
 Matches 20: Conservative 11: Mismatches 23: Indels 1: Gaps 1:

DB 108 FGAAGNRPLWIFKNDVIRLQKVMVNSUKITIGRGAKVEIIN 150  
 QY 67 FTLMGRRLPWIFSGNNKIKMKMYIAIKYKIDGRRAEVSVMVNSAKFIKRAVLCII 149

RESULT 8  
 ID PE56\_LYCES STANDARD: PRI: 398 AA.

AC P15721:  
 DT 01-APR-1990 (Rel. 14, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE PROBABLE PECTATE LYASE P56 (CURSOR (EC 4.2.2.2)).  
 GN LAT56.

OS Lycopersicon esculentum (Tomato)  
 OC Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta:  
 OC euphyllophytes: Spermatophyta: Magnoliophyta: eudicotyledons:  
 OC core eudicots: Asteridae: euasterids I: Solanales: Solanaceae:  
 OC Solanum.

```

PN 111
PP SEQUENCE FROM N.A.
RC STRAIN-CV VF36; TISSUE=ANTHER;
RX MEDLINE: 9122485.
RA Wase R.A., Yarauchi J., Larabell S.K., Usin V.M., McCormick S.,
RI "Molecular and genetic characterization of two pollen-expressed genes
RI that have sequence similarity to pectate lyases of the plant pathogen
RI Erwinia."
RL Plant Mol Biol 14:17-28(1990).
RN [2]
RV REVIEWS.
RC STRAIN-CV VF36; TISSUE=ANTHER;
RX Wase R.A.;
RI Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.
RI FUNCTION: MIGHT BE NEEDED DURING POLLEN DEVELOPMENT AND TUBE
RI GROWTH.
CC -!- CATALYTIC ACTIVITY: ELIMINATIVE CLEAVAGE OF PECTATE TO GIVE
CC OLIGOSACCHARIDES WITH 4-DEOXY-ALPHA-D-MANN-4-ENURONOSYL GROUPS
CC AT THEIR NON-REDUCING ENDS.
CC -!- TISSUE SPECIFICITY: EXPRESSED IN ANTHERS AND POLLEN.
CC -!- SIMILARITY: BELONGS TO THE POLYSACCHARIDE LYASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed, usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: X15500; CAA33524.1;
CC PIR: S05383; SC8383;
CC PFAM: PF00544; PectLyase;
CC PRINTS: PR00807; AMBALLERGEN.
CC Lyase: Multimeric family: Signal.
CC SIGNAL: 1 27
CC CHAIN: 28 398
CC ACT_SITE: 273 273
CC CARBOHYD: 135 135
CC POTENTIAL:
CC POTENTIAL:
CC SEQUENCE 398 AA: 44563 MW: 80676250.6 IBC708 CRC64:
Query Match 13.5% Score 138; DB 1; Length 398;
Best Local Similarity 35.5% Pred. No 1;37e-07;
Matches 22; Conservative 17; Mismatches 5; Indels 5; Gaps 4;

DB 106 FGVIQPKWITFARSMRLTLRELIVSSNKIDGR-GK--YVH-ANGAGIKISASNV 162
QY 67 FLMGRFLWILFSGNNMKIKKMWYIAGYKTFDGRRAEVSVHV-NGAKF-IRRVGGII 124

DE 163 15 164
QY 125 AA 126

RESULT 9
ID WP12 AMBAR STANDARD: PRI: 398 AA.
AC P27760;
DI 01-AUG-1992 (Rel. 23, Created)
DI 15-DEC-1998 (Rel. 37, Last sequence update)
DE POLLEN ALLERGEN AMB A 1.2 PRECURSOR (ANTIGEN E).
OS Ambrosia artemisiifolia (Short ragweed).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Asteridae; euasterids II; Asterales; Asteraceae;
OC Asteroideae; Heliantheae; Ambrosia.
RN [-]
RV REVIEWS.
RC STRAIN-CV VF36; TISSUE=ANTHER;
RX MEDLINE: 91093235.
RA Rafnar T., Griffith I.J., Kuo M.-C., Bond J.F., Rogers B.L.,
RA Klapper D.G.;

```

```

RT "Cloning of Amb a 1 (antigen E), the major allergen family of short
RT ragweed pollen."
RL J. Biol. Chem. 266:1229-1236(1991).
RN [2]
RP SEQUENCE FROM N.A. AND VARIANTS.
RC TISSUE=POLLEN;
RX MEDLINE: 92234570.
RA Griffith I.J., Pollack J., Klapper D.G., Rogers B.L., Nault A.M.;
RI "Sequence polymorphism of Amb a 1 and Amb a 11, the major allergens
RI in Ambrosia artemisiifolia (Short ragweed)."
RL Int. Arch. Allergy Appl. Imm Biol 96:296-304(1991).
RN [2]
RV SUBUNIT: MONOMER.
CC -!- TISSUE SPECIFICITY: POLLEN AND FLOWERS.
CC -!- DISEASE: ONE OF THE MAJOR ALLERGENS OF THE RAGWEED POLLEN.
CC -!- SIMILARITY: BELONGS TO THE POLYSACCHARIDE LYASE FAMILY.
CC AMB A 1/AMB A 11/CRY 3.1 SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed, usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: M52981; AAA32656.1;
CC EMBL: M80559; AAA32667.1;
CC PIR: B39099; B39099;
CC PFAM: PF00544; PectLyase;
CC PRINTS: PR00807; AMBALLERGEN.
CC Antigen: Allergen; Signal: Multimeric family: Polymorphic.
CC SIGNAL: 1 25
CC CHAIN: 26 398
CC MOD_RES: 226 226
CC VARIANT: 345 345
CC VARIANT: 381 381
CC SEQUENCE 398 AA: 43664 MW: 92006620.9 BT4150 CRC54:
Query Match 13.4% Score 137; DB 1; Length 398;
Best Local Similarity 39.7% Pred. No 1.37e-07;
Matches 23; Conservative 19; Mismatches 23; Indels 2; Gaps 2;

DB 109 FFAAQNRLPKLIFKKNVHLNDELAVNSDKTIDGRGVKVNVA-GLILNN-VKNI 164
QY 67 FLMGRFLWILFSGNNMKIKKMWYIAGYKTFDGRRAEVSVHV-NGAKF-IRRVGGII 124

RESULT 10
ID 9612 LYCES STANDARD: PRI: 474 AA.
AC P24396;
DI 01-MAR-1992 (Rel. 21, Created)
DI 01-MAR-1992 (Rel. 21, Last sequence update)
DI 15-JUL-1999 (Rel. 38, Last annotation update)
DE STYLE DEVELOPMENT-SPECIFIC PROTEIN 9612 PRECURSOR.
GN 9612.
OS Lycopersicon esculentum (Tomato).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Asteridae; euasterids I; Solanales; Solanaceae;
OC Solanum.
RN [1]
RV REVIEWS.
RC STRAIN-CV VF36; TISSUE=PISTIL;
RX MEDLINE: 91117185.
RA Badellier K.A., Smith A.G., Gasser C.S.;
RI "Regulation of a stylar transmitting tissue-specific gene in
RI wild-type and transgenic tomato and tobacco."
RL Mol. Gen. Genet. 224:183-192(1990).
CC -!- FUNCTION: MAY HAVE A ROLE IN THE DEVELOPMENT OF THE TRANSMITTING
CC TISSUE OF THE STYLE AND/OR IN THE EVENTS RELATED TO POLLINATION
CC SUCH AS SOME ASPECT IN THE FACILITATION OF COMPATIBLE POLLEN TUBE
CC GROWTH.
CC -!- SUBCELLULAR LOCATION: SECRETED.

```

CC 1- TISSUE SPECIFICITY: PREDOMINANTLY FOUND IN THE PISTIL WHERE IT IS  
 CC FOUND IN THE OUTER FIVE LAYERS OF THE STRANDS OF TRANSMITTING  
 CC TISSUE WITHIN THE UPPER TWO-THIRDS OF THE STYLE. FOUND AT MUCH  
 CC LOWER LEVELS IN THE ANthers AND VEGETATIVE ORGANS.  
 CC 1- DEVELOPMENTAL STAGE: MAXIMUM LEVELS ARE FOUND DURING ANTHESIS.  
 CC 1- SIMILARITY: 54% IDENTICAL TO TOMATO PROTEINS P59 (AC P.5722)  
 CC AND P56 (AC P.5722).  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: X55193; CAA36979.1;  
 CC PIR: S12209; S12209;  
 CC PFAM: PF00544; pec\_lyase; 1;  
 CC PRINTS: PROC807; AMBALLERGEN;  
 CC SIGNAL: 1 20 POTENTIAL;  
 CC CHAIN 21 404 STYLE DEVELOPMENT-SPECIFIC PROTEIN 9612;  
 CC CARBOHYD 37 37 POTENTIAL;  
 CC CARBOHYD 191 191 POTENTIAL;  
 CC SEQUENCE 434 AA: 44298 MW: 826563512608675 CRC64;  
 CC  
 CC Query Match 12.48; Score 137; DB 1; Length 404;  
 CC Best Local Similarity 54.58; Pred. No. 1370-87;  
 CC Matches 18; Conservative 5; Mismatches 10; Indels 0; Gaps 0;  
 CC  
 CC 120 RPLWTFQKQVIOIKQELVMSNYKIFGRGASV 152  
 CC  
 CC 74 PLNIFSGNNKTKLMPVYIAGYKTEGKRAEVS 106  
 CC  
 CC RESULT 11  
 CC ID MP22-AMBAR STANDARD; PRT: 397 AA;  
 CC AC 927762;  
 CC DT 01-AUG-1992 (Rel. 23, Created)  
 CC U1 01-AUG-1992 (Rel. 23, Last sequence update)  
 CC U2 01-DEC-1998 (Rel. 37, Last annotation update)  
 CC DE POLLEN ALLERGEN AMB A 2 PRECURSOR (ANTH: K) (ANTH:ZEN AMB A 11);  
 CC Ambrosia artemisiifolia (short ragweed);  
 CC Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta;  
 CC euphyllophytes: Spermatophyta: Magnoliophyta: eudicotyledons:  
 CC core eudicots: Asteridae: easterids 11: Asterales: Asteroideae:  
 CC Asteroideae: Euphorbiales: Ambrosia;  
 CC (1)  
 CC RY SEQUENCE FROM N.A.  
 CC RC TISSUE: FLOWER;  
 CC RX MEDLINE: 92234570;  
 CC RA Rogers B.L., Morgenstern J.P., Griffith I.J., Yu X.-B.,  
 CC RA Counsell C.M., Brauer A.W., King T.P., Garman R.D., Kuo M.-C. C.;  
 CC FT "Complete sequence of the allergen Amb alpha II. Recombinant  
 CC FT expression and reactivity with T cells from ragweed allergic  
 CC FT patients".  
 CC RL J. Immunol. 147:2547-2552(1991).  
 CC RN (2).  
 CC RP SEQUENCE FROM N.A., AND VARIANTS.  
 CC RC TISSUE: POLLEN;  
 CC RX MEDLINE: 92234570;  
 CC RA Griffith I.J., Pollock J., Klapper D.G., Rogers B.L., Nault A.K.,  
 CC RT "Sequence polymorphism of Amb a 1 and Amb a 11, the major allergens  
 CC RT in Ambrosia artemisiifolia (short ragweed)".  
 CC RL Int. Arch. Allergy Appl. Immunol. 96:296-304(1991).  
 CC CC 1- SUBUNIT: MONOMER;  
 CC 1- TISSUE SPECIFICITY: POLLEN AND FLOWERS;  
 CC 1- DISEASE: ONE OF THE MAJOR ALLERGENS OF THE RAGWEED POLLEN;  
 CC 1- SIMILARITY: BELONGS TO THE POLYSACCHARIDE LYASE FAMILY 1.  
 CC AMB A 1/AMB A 11/CRY J 1 SUBFAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL Outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: M80561; AAA32671.1;  
 CC PIR: A45459; A45459;  
 CC PFAM: PF00544; pec\_lyase; 1;  
 CC PRINTS: PROC807; AMBALLERGEN;  
 CC Antigen: Allergen: Signal: Polymorphism;  
 CC SIGNAL: 26 256 POTENTIAL;  
 CC CHAIN 26 256 POLLEN ALLERGEN AMB A 2;  
 CC MOD\_RES 726 726 BLOCKED;  
 CC VARIANT 70 70 N 20 C (DETECTED ONLY IN FLOWER (NA));  
 CC VARIANT 138 138 K 20 Y (DETECTED ONLY IN FLOWER (NA));  
 CC VARIANT 321 321 K 20 P (DETECTED ONLY IN FLOWER (NA));  
 CC SEQUENCE 397 AA: 44082 MW: 82617840410108064;  
 CC  
 CC Query Match 12.88; Score 141; DB 1; Length 473;  
 CC Best Local Similarity 42.38; Pred. No. 1370-86;  
 CC Matches 22; Conservative 8; Mismatches 40; Indels 0; Gaps 2;  
 CC  
 CC 114 RPLWTFQKQVIOIKQELVMSNYKIFGRGASV 152  
 CC  
 CC 73 RPLWTFQKQVIOIKQELVMSNYKIFGRGASV 106  
 CC  
 CC RESULT 12  
 CC ID P559-LYCES STANDARD; PRT: 445 AA;  
 CC AC P15722;  
 CC DT 01-APR-1990 (Rel. 14, Created)  
 CC U1 01-APR-1990 (Rel. 14, Last sequence update)  
 CC U2 15-JUL-1999 (Rel. 38, Last annotation update)  
 CC DE PROBABLE PECTATE LYASE P59 PRECURSOR (P. 4.2.2.2)  
 CC LA159;  
 CC GN Lycopersicon esculentum (Tomato)  
 CC OS Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta;  
 CC euphyllophytes: Spermatophyta: Magnoliophyta: eudicotyledons:  
 CC core eudicots: Asteridae: easterids 11: Solanales: Solanales;  
 CC Solanum;  
 CC RN 1;  
 CC RC SEQUENCE FROM N.A.  
 CC RP STRAIN: CV. VF35; TISSUE: ANTH: K;  
 CC PX MEDLINE: 92322485;  
 CC RA Wang P.A., Yamamoto T.;  
 CC FT "Molecular and genetic characterization of the gene encoding pectate  
 CC FT lyase that have sequence similarity to pectate lyase 1 from potato  
 CC FT tubers".  
 CC RL Plant Mol. Biol. 34:17-26(1990).  
 CC CC 1- FUNCTION: MIGHT BE NEEDED DURING POLLEN DEVELOPMENT AND FERTI-  
 CC CC GROWTH.  
 CC 1- CATALYTIC ACTIVITY: ELIMINATIVE CLEAVAGE OF PECTATE TO GIVE  
 CC CC OLIGOSACCHARIDES WITH 4-DEOXY-ALPHA-D-MANN-4-ENURONOSYL GROUPS  
 CC AT THEIR NON-REDUCING ENDS.  
 CC 1- TISSUE SPECIFICITY: EXPRESSED IN ANthers AND POLLEN.  
 CC 1- SIMILARITY: BELONGS TO THE POLYSACCHARIDE LYASE FAMILY 1.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: X15499; CAA33523.1;  
 CC PIR: S27098; S27098;  
 CC PFAM: PF00544; pec\_lyase; 1;  
 CC PRINTS: PROC807; AMBALLERGEN;  
 CC Lyase: Multigene family: Signal;  
 CC SIGNAL: 1 22 POTENTIAL;





```
ID RRP2-RVEVZ STANDARD: PRI: 2149 AA.
AC P27316.
DI 01-AUG-1992 (rel. 21. Created:
DI 01-AUG-1992 (rel. 21. Last sequence update)
DI 01-AUG-1992 (rel. 21. Last annotation update)
DE RNA-DIRECTED RNA POLYMERASE (EC 2.7.7.45) (L PROTEIN).
GN Rift valley fever virus (strain ZH-548 M2) (RVFV).
NC Viruses: ssRNA negative-strand viruses; Bunyaviridae; Phlebovirus.
FN [1]
FX MEDLINE: 92020294.
SA Muller A., Arnetini C., Bouloy M., Prehaud C., Bishop D.H.L.
RI "Completion of the genome sequence of Rift valley fever phlebovirus
RI indicates that the L RNA is negative sense and codes for a putative
RI transcriptase-replicase."
RL Nucleic Acids Res. 19:5433-5437(1991).
CC -- CATALYTIC ACTIVITY: N NUCLEOSIDE TRIPHOSPHATE + N PYROPHOSPHATE
CC + RNA(N).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/annouce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X56444; CAA39836.1; -.
DI PIR: S18576; S18576
KW Transcriptase; Nucleosidylyltransferase; RNA-directed RNA polymerase.
SQ SEQUENCE 2149 AA; 243569 MW; 6D57395C79A58D7 CRC64;
Query Match: 9.6%; Score 99; DB 1; Length 2149;
Best Local Similarity 34.4%; Pred. No. 1,160-01;
Matches 22; Conservative 14; Mismatches 22; Indels 4; Gaps 4;
De 986 TSDDAPKNAAGSHVTKFPAIMCEFTSPKAWPL-ITGDSMTKKMMNNLN-YLKILNGH 1045
QY 1 1111 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
QY 45 TISTGPKKNNPWLQQA-KLQCFIMGRRLWITFGNNMNIKLKMPYIASY-KTIDGR 102
Gn 1046 R 1046
QY 103 R 103
Search completed: Mon Jun 19 15:57:01 2000
Job time : 12 secs.
```

.....  
M E B E R S H I P  
.....  
(TM)  
.....

Kellogg 3-1A John F. Collins, Biocomputing Research Unit,  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd  
Mesich\_pf protein - protein database search, using Smith-Waterman algorithm  
Run on: Mon Jun 19 15:57:18 2000; MasPar time 18.25 Seconds  
Tabular output not generated. 509,207 Million cell updates/sec  
File: US-09-142-524A-3  
Description: (1-134) from US09142524A.pep  
Perfect Score: 1026  
Sequence: 1 MKVIVANQGFENRRVFIKR.....KEIRRVGGIIIAAYQNPAWK 134

Scoring table: PAM 150  
Gap 11  
Searched: 225878 seqs, 69334122 residues  
Post-processing: Minimum Match 0%  
Listing first 45 summaries  
Database: spiremb12  
1:SP\_dichaea 2:sp\_bacteria 3:sp\_fungi 4:sp\_human  
5:sp\_invertebrate 6:sp\_mammal 7:sp\_mhc 8:sp\_organella  
9:sp\_phase 10:sp\_plant 11:sp\_protein 12:sp\_unclassified  
13:sp\_vertebrate 14:sp\_virus  
Statistics: Mean 42.950; Variance 75.878; scale 0.566  
Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description | Pred. No.              |
|------------|-------|-------------|--------|-------|-------------|------------------------|
| 1          | 224   | 21.8        | 367    | 10    | Q92N07      | POLLEN MAJOR ALLERGEN  |
| 2          | 210   | 20.5        | 375    | 10    | Q96385      | CHAOI PRECURSOR        |
| 3          | 150   | 14.6        | 181    | 10    | Q23666      | PUTATIVE PECTATE LYASE |
| 4          | 148   | 14.4        | 431    | 10    | Q23017      | TIG11.7 PROTEIN        |
| 5          | 142   | 13.7        | 455    | 10    | Q64510      | PUTATIVE PECTATE LYASE |
| 6          | 140   | 13.6        | 390    | 10    | Q65388      | FL2F1.22 PROTEIN       |
| 7          | 138   | 13.5        | 459    | 10    | Q23665      | PUTATIVE PECTATE LYASE |
| 8          | 136   | 13.3        | 394    | 10    | Q65456      | PECTATE LYASE LIKE PRC |
| 9          | 134   | 13.1        | 384    | 10    | Q65457      | PECTATE LYASE LIKE PRC |
| 10         | 134   | 13.1        | 425    | 10    | Q24416      | PECTATE LYASE          |
| 11         | 133   | 13.0        | 226    | 10    | Q23667      | PUTATIVE PECTATE LYASE |
| 12         | 131   | 12.8        | 438    | 10    | Q43862      | POLLEN SPECIFIC PECTAT |
| 13         | 130   | 12.7        | 450    | 10    | Q40319      | PECTATE LYASE HOMOLOG  |
| 14         | 129   | 12.6        | 227    | 10    | Q23668      | PUTATIVE PECTATE LYASE |
| 15         | 124   | 12.1        | 274    | 10    | Q24159      | PUTATIVE PECTATE LYASE |
| 16         | 122   | 11.9        | 401    | 10    | Q24554      | PECTATE LYASE PRECURSO |
| 17         | 118   | 11.5        | 398    | 10    | Q43783      | PECTATE LYASE (EC 4.2. |
| 18         | 113   | 10.7        | 4450   | 10    | Q44928      | GRAMICIDIN'S SYNTHETAS |
| 19         | 103   | 10.0        | 1484   | 10    | Q9X585      | POLYPROTEIN            |
| 20         | 100   | 9.7         | 238    | 10    | Q02901      | HYPOTHETICAL 27.5 KD P |

|    |     |     |      |    |        |                              |          |  |  |
|----|-----|-----|------|----|--------|------------------------------|----------|--|--|
| 21 | 100 | 9.7 | 333  | 10 | OB1494 | F9D12.7 PROTEIN              | 1,048-01 |  |  |
| 22 | 98  | 9.6 | 626  | 5  | Q46039 | EG1103B4.3 PROTEIN           | 2,048-01 |  |  |
| 23 | 96  | 9.4 | 190  | 2  | Q9X000 | HYPOTHETICAL 22.4 KD P       | 3,978-01 |  |  |
| 24 | 95  | 9.3 | 108  | 14 | Q09783 | CAPSIN                       | 5,548-01 |  |  |
| 25 | 95  | 9.3 | 3229 | 5  | Q26912 | PROTEIN 1 OF A LIPID         | 6,548-01 |  |  |
| 26 | 94  | 9.2 | 564  | 2  | F22553 | HYPOTHETICAL 53.9 KD P       | 7,048-01 |  |  |
| 27 | 94  | 9.2 | 815  | 5  | Q23850 | VACUOLAR PROTON ATPASE       | 7,048-01 |  |  |
| 28 | 93  | 9.1 | 875  | 11 | F97475 | ALKALINE PHOSPHATASE         | 1,048-00 |  |  |
| 29 | 93  | 9.1 | 875  | 11 | F17476 | ALKALINE PHOSPHATASE         | 1,048-00 |  |  |
| 30 | 93  | 9.1 | 875  | 11 | G14390 | PH13-6 ANTIGEN               | 1,048-00 |  |  |
| 31 | 93  | 9.1 | 876  | 11 | F70541 | PHOSPHOGLUCOSYLTRANSFERASE 1 | 1,048-00 |  |  |
| 32 | 93  | 9.1 | 1347 | 2  | Q38426 | XILANASE                     | 1,048-00 |  |  |
| 33 | 92  | 9.0 | 215  | 3  | Q96095 | A2 MATING TYPE 1             | 1,048-00 |  |  |
| 34 | 92  | 9.0 | 360  | 2  | Q23639 | HYPOTHETICAL 41.9 KD P       | 1,048-00 |  |  |
| 35 | 91  | 8.9 | 226  | 2  | Q56807 | HYPOTHETICAL 25.5 KD P       | 2,018-00 |  |  |
| 36 | 91  | 8.9 | 426  | 10 | Q80147 | S GLYCOPROTEIN 1             | 2,018-00 |  |  |
| 37 | 91  | 8.9 | 438  | 3  | Q13741 | PUTATIVE RNA-BINDING P       | 2,018-00 |  |  |
| 38 | 91  | 8.9 | 596  | 10 | Q49550 | 1-PROSPHATIDYLIN-SITIN       | 2,018-00 |  |  |
| 39 | 91  | 8.9 | 652  | 14 | Q9MLD7 | ENVELOPE GLYCOPROTEIN        | 2,018-00 |  |  |
| 40 | 90  | 8.8 | 56   | 1  | Q29714 | PROTEIN TRANSFERASE S        | 2,768-00 |  |  |
| 41 | 90  | 8.8 | 146  | 11 | Q88982 | EXTRACELLULAR CALCIUM        | 2,768-00 |  |  |
| 42 | 90  | 8.8 | 377  | 2  | P71454 | D-ALA-D-ALA LIGASE           | 2,768-00 |  |  |
| 43 | 90  | 8.8 | 450  | 10 | Q40844 | VICILIN-LIKE STORAGE P       | 2,768-00 |  |  |
| 44 | 90  | 8.8 | 589  | 1  | Q53723 | DNA LIGASE                   | 2,768-00 |  |  |
| 45 | 89  | 8.7 | 300  | 14 | Q65581 | ULF POLYPEPTIDE              | 3,758-00 |  |  |

ALIGNMENTS

RESULT 1  
ID Q92N07 PRELIMINARY: PRT: 367 AA.  
AC Q92N07  
DT 01-MAY-1999 (Tremblrel. 10, Created)  
DI 01-MAY-1999 (Tremblrel. 10, Last sequence update)  
DI 01-NOV-1999 (Tremblrel. 12, Last annotation update)  
DE POLLEN MAJOR ALLERGEN 1-2  
OS Juniperus ashei (Oxark white cedar)  
OC Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta  
OC euphyllophytes: Spermatophyta: Coniferopsida: Coniferales  
OC Taxodiaceae: Juniperus  
RN 11  
RP SEQUENCE FROM N.A.  
RA MIDORO-HOSIUTI T.M., GOLDBL. A.M., KOSKOV A., WOOD T.S.,  
RA BROOKS E.G.  
RT "Molecular cloning of mount cedar (Juniperus ashei) pollen major  
allergen, Jun a 1.1;"  
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AF106563; AAO3609.1  
DR EMBL: AF106562; AAO3608.1  
DR EMBL: AF106561; AAO3607.1  
DR MENDEL: 35544; Junas:1288,36544  
DR MENDEL: 35545; Junas:1288,36545  
SQ SEQUENCE 367 AA: 35924 MW: 42246.4 251342

Query Match 21.8% Score 224; DB 10; Length 367;  
Best Local Similarity 55.8%; Pred. No. 3,598-24;  
Matches 29; Conservative 10; Mismatches 13; Indels 0; Gaps 0;

DB 87 KALWIFSONNKLKMLPYVAGHKITDGGADVHLGGPGCLPFMKVSRVI 138  
OY 73 RLWIFSONNKLKMLPYVAGHKITDGGADVHLGGPGCLPFMKVSRVI 124

RESULT 2  
ID Q96385 PRELIMINARY: PRT: 375 AA.  
AC Q96385  
DT 01-FEB-1997 (Tremblrel. 02, Created)  
DI 01-FEB-1997 (Tremblrel. 0, Last sequence update)  
DI 01-NOV-1999 (Tremblrel. 12, Last annotation update)  
DE CHAOI PRECURSOR  
OS Chamaecyparis obtusa  
OC Chamaecyparidaceae: Streptophyta: Embryophyta: Tracheophyta  
OC Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta  
OC euphyllophytes: Spermatophyta: Coniferopsida: Coniferales  
OC Taxodiaceae: Chamaecyparis



DE 166 RPLWIFARSMILKQGLLIINTKTIIDGRGAKI-YITGGAGLTQFVNV 215  
 QY 73 RPLWIFSGNMNKLKMPYAGYKTFDGRRAEVSYVHNGAKF-IRRVCGII 124

RESULT 6 PRELIMINARY: PRT: 390 AA.  
 ID 065456  
 AC 065456  
 DE 01-AUG-1998 (TRENBLER: 07, Created)  
 DT 01-AUG-1998 (TRENBLER: 07, Last sequence update)  
 DT 01-NOV-1999 (TRENBLER: 12, Last annotation update)  
 DE PEPTATE LYASE LIKE PROTEIN  
 GN F120.180  
 OS Arabidopsis thaliana (Mouse-ear cress)  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;  
 OC core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;  
 OC Arabidopsis.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA BEVAN M., WEDLER H., WAMBUT P., BANCROFT I., MEYER H.W., MAYER K.,  
 RA SCHUELLER C.,  
 RA Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA EU ARABIDOPSIS SEQUENCING PROJECT;  
 RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AL022140; CAA18112.1;  
 DR MENDEL: 29068; Atath1088;29068.  
 DR PFAM: PF00544; pec\_lyase; 1.  
 DR PRINTS: PRO0807; AMBALLERGEN.  
 KW Lyase.  
 SQ SEQUENCE 390 AA: 43354 MW: 8332484 CRC32:  
 Query Match 13.5% Score 140: DB 10: Length 390:  
 Best Local Similarity 36.4% Pred. No. 3.27e-08:  
 Matches 16: Conservative 15: Mismatches 11: Indels 2: Gaps 1:

DE 95 YAVLSKPKWTFARDMVIVLRNELIMNSYKTIIDGRGAKVEIAY 138  
 QY 67 FTLMGRPRVWIFSGNMNKLKMPYAGYKTFDGR--RAEVSY 108

RESULT 7 PRELIMINARY: PRT: 450 AA.  
 ID 023665  
 AC 023665  
 DE 01-JAN-1998 (TRENBLER: 05, Created)  
 DT 01-JAN-1998 (TRENBLER: 05, Last sequence update)  
 DT 01-NOV-1999 (TRENBLER: 12, Last annotation update)  
 DE PUTATIVE PEPTATE LYASE  
 GN 59.  
 OS Arabidopsis thaliana (Mouse-ear cress)  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;  
 OC core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;  
 OC Arabidopsis.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA KULIKAVSKAS R., MCCORMICK S.,  
 RA Identification of the tobacco and Arabidopsis homologues of the  
 RI pollen-expressed LA-59 gene of tomato.  
 RL Plant Mol. Biol. 34:809-814(1997).  
 DR EMBL: U83619; AAB694759.1;  
 DR MENDEL: 25559; Atath1089;25559.  
 DR PFAM: PF00544; pec\_lyase; 1.  
 DR PRINTS: PRO0807; AMBALLERGEN.  
 KW Lyase.  
 SQ SEQUENCE 459 AA: 51420 MW: 4185988 CRC32:  
 Query Match 13.5% Score 138: DB 10: Length 459:  
 Best Local Similarity 44.2% Pred. No. 7.30e-08:  
 Matches 23: Conservative 14: Mismatches 12: Indels 3: Gaps 3:

DE 171 PLWIFARSMILKQGLLIINTKTIIDGRGARY-YI-MEGAGLTQFVNV 220

QY 74 PLWIFSGNMNKLKMPYAGYKTFDGRRAEVSYVHNGAKF-IRRVCGII 124  
 RESULT 8 PRELIMINARY: PRT: 394 AA.  
 ID 065456  
 AC 065456  
 DE 01-AUG-1998 (TRENBLER: 07, Created)  
 DT 01-AUG-1998 (TRENBLER: 07, Last sequence update)  
 DT 01-NOV-1999 (TRENBLER: 12, Last annotation update)  
 DE PEPTATE LYASE LIKE PROTEIN  
 GN F120.180  
 OS Arabidopsis thaliana (Mouse-ear cress)  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;  
 OC core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;  
 OC Arabidopsis.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA BEVAN M., WEDLER H., WAMBUT P., BANCROFT I., MEYER H.W., MAYER K.,  
 RA SCHUELLER C.,  
 RA Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA EU ARABIDOPSIS SEQUENCING PROJECT;  
 RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AL022140; CAA18112.1;  
 DR MENDEL: 29068; Atath1088;29068.  
 DR PFAM: PF00544; pec\_lyase; 1.  
 DR PRINTS: PRO0807; AMBALLERGEN.  
 KW Lyase.  
 SQ SEQUENCE 394 AA: 43476 MW: 96396175 CRC32:  
 Query Match 13.3% Score 136: DB 10: Length 394:  
 Best Local Similarity 36.4% Pred. No. 1.62e-07:  
 Matches 16: Conservative 15: Mismatches 13: Indels 2: Gaps 1:

DE 105 YGVICAKPKWTFARDMVIVLRNELIMNSYKTIIDGRGAKVEIAY 149  
 QY 67 FTLMGRPRVWIFSGNMNKLKMPYAGYKTFDGR--RAEVSY 108

RESULT 9 PRELIMINARY: PRT: 394 AA.  
 ID 065457  
 AC 065457  
 DE 01-AUG-1998 (TRENBLER: 07, Created)  
 DT 01-AUG-1998 (TRENBLER: 07, Last sequence update)  
 DT 01-NOV-1999 (TRENBLER: 12, Last annotation update)  
 DE PEPTATE LYASE LIKE PROTEIN  
 GN F120.180  
 OS Arabidopsis thaliana (Mouse-ear cress)  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;  
 OC core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;  
 OC Arabidopsis.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA BEVAN M., WEDLER H., WAMBUT P., BANCROFT I., MEYER H.W., MAYER K.,  
 RA SCHUELLER C.,  
 RA Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA EU ARABIDOPSIS SEQUENCING PROJECT;  
 RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AL022140; CAA18112.1;  
 DR MENDEL: 29068; Atath1088;29068.  
 DR PFAM: PF00544; pec\_lyase; 1.  
 DR PRINTS: PRO0807; AMBALLERGEN.  
 KW Lyase.  
 SQ SEQUENCE 394 AA: 43299 MW: 88949309 CRC32:  
 Query Match 13.1% Score 134: DB 10: Length 394:  
 Best Local Similarity 36.4% Pred. No. 3.58e-07:

Matches 16: Conservative 13: Mismatches 13: Indels 2: Gaps 1:

DB 105 NVVICAKPKITFAKUMVITLANELMKNVSKYTIQSI AKVEIAY 148  
 QY 67 FLMRPHHPIWITFSNNMKIKMPMYIAGYATFQI -RAEVS 108

RESULT 13  
 ID Q24416 PRELIMINARY: PRI: 405 AA.  
 AC Q24416  
 DT 01-JAN-1998 (FEMBLrel: 05, Created)  
 DI 01-JAN-1998 (FEMBLrel: 05, Last sequence update)  
 DE 01-NOV-1999 (FEMBLrel: 12, Last annotation update)  
 DE PUTATIVE PECTATE LYASE  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 CC euphyllophytes; Spermatophyta; Magnoliopsida; eudicotyledons;  
 CC core eudicots; Rosidae; eurosids I; Rosales; Rosaceae; Fragaria.  
 CC  
 RN  
 RP SEQUENCE FROM N.A.  
 RC STRAIN: CHANDLER.  
 RX MEDLINE: 9435972.  
 RA MEDINA ESCOBAR N., CAPERNAS J., MOYANO E., CABALLERO J.L.,  
 RA MENDEL-ALANIZ J.  
 RT "Cloning, characterization, and expression pattern of a  
 RT strawberry ripening-specific cDNA with sequence homology to pectate  
 RT lyase from higher plants."  
 RI Plant Mol Biol 34:867-877(1997).  
 DR EMBL: 063551; AAB71208.1.  
 DR MENDEL: 250597; Frax1068:45059.  
 DR PRAM: PF00544; pec\_lyase:1  
 KW Lyase.  
 SQ SEQUENCE 405 AA: 45744 MW: 289944AC CRC32:

Query Match 13.1% Score 134: DB 10: Length 405.  
 Best Local Similarity 36.5% Pred. No. 3,590-07:  
 Matches 19: Conservative 13: Mismatches 20: Indels 1: Gaps 1:

DB 166 RPLWIKRQMTITLQELIMNSFKIDARQNVV HAYGSCILIFVIVNI 216  
 QY 73 RPLWIFSGNNKIKMPMYIAGYATFQIRAEVSIVHNSAKPIRWGII 124

RESULT 13  
 ID Q23657 PRELIMINARY: PRI: 226 AA.  
 AC Q23657  
 DT 01-JAN-1998 (FEMBLrel: 05, Created)  
 DI 01-JAN-1998 (FEMBLrel: 05, Last sequence update)  
 DE 01-NOV-1999 (FEMBLrel: 12, Last annotation update)  
 DE PUTATIVE PECTATE LYASE (FRAGMENT).  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 CC euphyllophytes; Spermatophyta; Magnoliopsida; eudicotyledons;  
 CC core eudicots; Rosidae; eurosids I; Brassicales; Brassicaceae;  
 CC Arabidopsis.  
 RN  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE: 97422403.  
 RA KULIKAVSKAS R., MCCORMACK S.J.  
 RT "Identification of the tobacco and Arabidopsis homologues of the  
 RT pollen-expressed LAYS9 gene of tomato."  
 RI Plant Mol Biol 34:809-814(1997).  
 DR EMBL: 083521; AAB69761.1.  
 DR MENDEL: 25561; Arab1088:25561.  
 DR PRAM: PF00544; pec\_lyase:1.  
 KW Lyase.  
 SQ SEQUENCE 226 AA: 25103 MW: 5F1C8DD2 CRC32:

Query Match 13.0% Score 133: DB 10: Length 226:

Best Local Similarity 51.5% Pred. No. 5,320-07:  
 Matches 17: Conservative 6: Mismatches 10: Indels 0: Gaps 0:

DB 54 PLWLIIFURQNTIOLKEELIMNSEKILQHGASV 86  
 QY 74 PLWLIIFSGNNKIKMPMYIAGYATFQIRAEV 106

RESULT 12  
 ID Q43822 PRELIMINARY: PRI: 438 AA.  
 AC Q43822  
 DT 01-NOV-1996 (FEMBLrel: 01, Created)  
 DI 01-NOV-1996 (FEMBLrel: 01, Last sequence update)  
 DE 01-NOV-1999 (FEMBLrel: 12, Last annotation update)  
 DE POLLEN SPECIFIC PECTATE LYASE HOMOLOGUE (EM 4.2.1.2)  
 DE (PECTATE TRANSELMINASE).  
 OS Zea mays (Maize).  
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 CC euphyllophytes; Spermatophyta; Magnoliopsida; eudicotyledons;  
 CC Poaceae; Zea.  
 RN  
 RP SEQUENCE FROM N.A.  
 RC STRAIN: CV. GOLDEN BUSH.  
 RX MEDLINE: 9408358.  
 RA TURICH N.P., HAMILTON D.A., MASARENAS J.P.  
 RT "Isolation and characterization of pollen-specific cDNA from a  
 RT sequence homology to pectate lyase and pectate lyase."  
 RI Plant Mol Biol 23:1061-1065(1997).  
 CC  
 CC CATALYTIC ACTIVITY: EC:4.2.1.2: PECTATE 4-EPICARBOXYLASE  
 CC GLUCOSACCHARIDES WITH 4-DEOXY-ALPHA-D-GLUC-4-EN-1,3-DIOL  
 CC THEIR NON-REDUCING ENDS.  
 DR EMBL: E20140; AAA16476.1.  
 DR MENDEL: 11165; Zeama1059:11165.  
 DR PRAM: PF00544; pec\_lyase:1.  
 DR PRINTS: PR00807; AMBALLERGEN.  
 KW Lyase.  
 SQ SEQUENCE 438 AA: 49148 MW: 1585369 CRC32:

Query Match 12.8% Score 131: DB 10: Length 438.  
 Best Local Similarity 47.1% Pred. No. 1,760-11:  
 Matches 16: Conservative 11: Mismatches 10: Indels 0: Gaps 0:

DB 156 RPLKIVFAHQVLEIQLGLIVNNSKIDGRQAV 189  
 QY 73 RPLWIFSGNNKIKMPMYIAGYATFQIRAEV 106

RESULT 13  
 ID Q40319 PRELIMINARY: PRI: 450 AA.  
 AC Q40319  
 DT 01-NOV-1996 (FEMBLrel: 01, Created)  
 DI 01-NOV-1996 (FEMBLrel: 01, Last sequence update)  
 DE 01-NOV-1999 (FEMBLrel: 12, Last annotation update)  
 DE PUTATIVE PECTATE LYASE HOMOLOGUE  
 OS Medicago sativa (Alfalfa).  
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 CC euphyllophytes; Spermatophyta; Magnoliopsida; eudicotyledons;  
 CC core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionaceae;  
 CC Medicago.  
 RN  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE: 97422403.  
 RA WU Y., QIU X., DU S., ERICKSON L.J.  
 RT "Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases  
 RT EMBL: 011472; AAB52411.1.  
 DR MENDEL: 9042; MedSa1068:904.  
 DR PRAM: PF00544; pec\_lyase:1.  
 DR PRINTS: PR00807; AMBALLERGEN.  
 KW Lyase.  
 SQ SEQUENCE 450 AA: 50155 MW: B35CFE9E CRC32:

Query Match 12.7% Score 130: DB 10: Length 450:  
 Best Local Similarity 37.3% Pred. No. 1,720-06:  
 Matches 22: Conservative 15: Mismatches 19: Indels 3: Gaps 3:





Best Local Similarity 95.7% Pred. NO. 6,59e-42; Mismatches 1; Indexes 13; Gaps 1;  
Matches 90; Conservative 1; Mismatches 1; Indexes 13; Gaps 1;

Query Match 20.3% Score 208; DB 1; Length 40;  
Best Local Similarity 100.0% Pred. NO. 5.42e-39;  
Matches 25; Conservative 0; Mismatches 0; Indexes 0; Gaps 0;

DB 5 RPLWIFSGNNMKIKMPMYIAGYK 30  
Query Match 20.3% Score 208; DB 1; Length 40;  
Best Local Similarity 100.0% Pred. NO. 5.42e-39;  
Matches 25; Conservative 0; Mismatches 0; Indexes 0; Gaps 0;

DB 79 RPLWIFSGNNMKIKMPMYIAGYK 97  
Query Match 20.3% Score 208; DB 1; Length 40;  
Best Local Similarity 100.0% Pred. NO. 5.42e-39;  
Matches 25; Conservative 0; Mismatches 0; Indexes 0; Gaps 0;

RESULTS 5  
ID R45549 standard; Protein: 20 AA.  
AC R45549;  
DE 13-JUL-1994 (first entry)  
KW Cry 3: pollen allergen peptide cry-3  
KW Japanese cedar; detection: allergy; treatment: diagnosis;  
KW T cell epitope; sensitivity  
OS Cryptomeria japonica  
PN W09401562-A  
PD 20-JAN-1994  
PF 15-JAN-1993; U00139  
PR 10-JUL-1992; W0-J05566;  
PR 01-SEP-1992; US-918990;  
PA (IMMU-) IMMUNOLOGIC PHARM CORP  
PI Bond CP, Garman RJ, Griffith K M, Kato T, Ito T,  
CR WPI: 94-035056/04;  
PT Antigens derived from Japanese Cedar; Pollen allergen Cry 3;  
PT contain at least two T cell epitope(s), used to treat or  
PT allergy  
PS Claim 1: Fig 13: 137bp; English  
CC The sequence is that of an isolated peptide of the Japanese cedar  
CC pollen allergen Cry 3 (amino acids 71-90). The peptide  
CC can be used for the treatment and diagnosis of allergies associated  
CC with Japanese cedar pollen. It has enhanced therapeutic properties  
CC but reduced side effects compared to naturally occurring allergens.  
SQ Sequence 20 AA;

Query Match 15.4% Score 159; DB 1; Length 20;  
Best Local Similarity 100.0% Pred. NO. 1.13e-04;  
Matches 20; Conservative 0; Mismatches 0; Indexes 0; Gaps 0;

DB 1 RPSNNMKIKMPMYIAGYK 20  
Query Match 15.4% Score 159; DB 1; Length 20;  
Best Local Similarity 100.0% Pred. NO. 1.13e-04;  
Matches 20; Conservative 0; Mismatches 0; Indexes 0; Gaps 0;

DB 79 RPSNNMKIKMPMYIAGYK 97  
Query Match 15.4% Score 159; DB 1; Length 20;  
Best Local Similarity 100.0% Pred. NO. 1.13e-04;  
Matches 20; Conservative 0; Mismatches 0; Indexes 0; Gaps 0;

RESULTS 6  
ID R42498 standard; Protein: 20 AA.  
AC R42498;  
DE 15-APR-1996 (first entry)  
KW Cry 3: Japanese cedar; pollen allergen peptide fragment (Cry-3);  
KW Cry 3: Japanese cedar; pollen allergen; modified; drug production;  
KW allergy; Cryptomeria japonica  
OS Cryptomeria japonica  
PN W09527786-A1  
PD 19-OCT-1995  
PF 06-APR-1995; U04249;  
PR 08-APR-1994; US-226248;  
PR 06-DEC-1994; US-350225;  
PA (IMMU-) IMMUNOLOGIC PHARM CORP  
PI Chen X, Evans S, Franzen SM, Kio M, Powers SP;  
PT Shaked Z;  
PS WPI: 95-366391/47;  
PT Modified cryptomeria japonica (Cry 3) I peptide(s) - useful for  
PT treating allergy to Japanese cedar pollen allergen or  
PT immunologically cross reactive allergens  
PS Disclosures: Figure 2: 60pp; English  
CC Novel peptides of cry 3 have been modified as a part of a  
CC preformulation scheme to develop an optimised drug product for  
CC therapeutic treatment of humans suffering from allergy to Japanese  
CC cedar pollen allergen or an allergen which is immunologically cross



CC reactive with Japanese cedar pollen allergen. Such modified peptides possess certain characteristics which render them particularly suitable for drug product formulation. Peptide fragments of CRY 1, modified and unmodified, are given in 884591-884525; this peptide fragment corresponds to amino acids 71-90 of the allergen protein.

CC Sequence 21 AA:

Query Match 15.4% Score 154.08 E-16 Length 21

Best Local Similarity 100.0% E-16 Length 21

Matches 20 Conservative 0 Mismatches 0 Indels 0 Gaps 0

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

PS Claim 47: Page 28: Jopp: Eng. 11  
CC Synthetic peptides based on portions of cedar pollen allergens A (884586) and B (884587) were tested for their ability to activate cedar allergen-specific T-cells but not allergen-specific B-cells. Antibodies to peptides (884586-884587) were identified as B-cell epitopes. These peptides, plus subsequent peptides (884588-884590) for T-cell recognition, and homologous peptides (884591-884592) can be used as immunotherapeutic agents to treat or prevent cedar pollenosis, avoiding side-effects such as anaphylaxis.

CC Sequence 17 AA:

Query Match 13.5% Score 135.08 E-16 Length 17

Best Local Similarity 100.0% E-16 Length 17

Matches 17 Conservative 0 Mismatches 0 Indels 0 Gaps 0

CC 1 NRIWQFAKLTGFILMG 17

CC 1 NRIWQFAKLTGFILMG 17

CC 1 NRIWQFAKLTGFILMG 17

CC 1 NRIWQFAKLTGFILMG 17

CC 1 NRIWQFAKLTGFILMG 17

CC 1 NRIWQFAKLTGFILMG 17

CC 1 NRIWQFAKLTGFILMG 17

CC 1 NRIWQFAKLTGFILMG 17

CC 1 NRIWQFAKLTGFILMG 17

CC 1 NRIWQFAKLTGFILMG 17

CC 1 NRIWQFAKLTGFILMG 17

CC 1 NRIWQFAKLTGFILMG 17

CC 1 NRIWQFAKLTGFILMG 17

CC 1 NRIWQFAKLTGFILMG 17

CC 1 NRIWQFAKLTGFILMG 17

CC 1 NRIWQFAKLTGFILMG 17

CC 1 NRIWQFAKLTGFILMG 17

CC 1 NRIWQFAKLTGFILMG 17

CC 1 NRIWQFAKLTGFILMG 17

CC 1 NRIWQFAKLTGFILMG 17

CC 1 NRIWQFAKLTGFILMG 17

CC 1 NRIWQFAKLTGFILMG 17

CC 1 NRIWQFAKLTGFILMG 17

CC 1 NRIWQFAKLTGFILMG 17

CC 1 NRIWQFAKLTGFILMG 17

CC 1 NRIWQFAKLTGFILMG 17

CC 1 NRIWQFAKLTGFILMG 17

CC 1 NRIWQFAKLTGFILMG 17

CC 1 NRIWQFAKLTGFILMG 17

CC 1 NRIWQFAKLTGFILMG 17

CC 1 NRIWQFAKLTGFILMG 17

CC 1 NRIWQFAKLTGFILMG 17

CC 1 NRIWQFAKLTGFILMG 17

CC 1 NRIWQFAKLTGFILMG 17

CC 1 NRIWQFAKLTGFILMG 17

CC 1 NRIWQFAKLTGFILMG 17

CC 1 NRIWQFAKLTGFILMG 17

CC 1 NRIWQFAKLTGFILMG 17

CC 1 NRIWQFAKLTGFILMG 17

CC 1 NRIWQFAKLTGFILMG 17

CC 1 NRIWQFAKLTGFILMG 17

CC 1 NRIWQFAKLTGFILMG 17

CC 1 NRIWQFAKLTGFILMG 17

CC 1 NRIWQFAKLTGFILMG 17

CC 1 NRIWQFAKLTGFILMG 17

CC 1 NRIWQFAKLTGFILMG 17

CC 1 NRIWQFAKLTGFILMG 17

CC 1 NRIWQFAKLTGFILMG 17

CC 1 NRIWQFAKLTGFILMG 17

CC 1 NRIWQFAKLTGFILMG 17

CC 1 NRIWQFAKLTGFILMG 17

CC 1 NRIWQFAKLTGFILMG 17

CC 1 NRIWQFAKLTGFILMG 17

CC 1 NRIWQFAKLTGFILMG 17

CC 1 NRIWQFAKLTGFILMG 17

CC 1 NRIWQFAKLTGFILMG 17

CC 1 NRIWQFAKLTGFILMG 17

CC 1 NRIWQFAKLTGFILMG 17

CC 1 NRIWQFAKLTGFILMG 17

CC 1 NRIWQFAKLTGFILMG 17

CC 1 NRIWQFAKLTGFILMG 17

CC 1 NRIWQFAKLTGFILMG 17

CC 1 NRIWQFAKLTGFILMG 17

CC 1 NRIWQFAKLTGFILMG 17

CC 1 NRIWQFAKLTGFILMG 17

CC reactive with Japanese cedar pollen allergen. Such modified peptides possess certain characteristics which render them particularly suitable for drug product formulation. Peptide fragments of CRY 1, modified and unmodified, are given in 884591-884525; this peptide fragment corresponds to amino acids 71-90 of the allergen protein.

CC Sequence 21 AA:

Query Match 15.4% Score 154.08 E-16 Length 21

Best Local Similarity 100.0% E-16 Length 21

Matches 20 Conservative 0 Mismatches 0 Indels 0 Gaps 0

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC reactive with Japanese cedar pollen allergen. Such modified peptides possess certain characteristics which render them particularly suitable for drug product formulation. Peptide fragments of CRY 1, modified and unmodified, are given in 884591-884525; this peptide fragment corresponds to amino acids 71-90 of the allergen protein.

CC Sequence 21 AA:

Query Match 15.4% Score 154.08 E-16 Length 21

Best Local Similarity 100.0% E-16 Length 21

Matches 20 Conservative 0 Mismatches 0 Indels 0 Gaps 0

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

CC 1 DSGNNKIKEMMYIARYK 20

WQSRQL (TM)

Release: 1A John F. Collins, PhD, Patent Research Unit,  
Copyright (c) 1994-1998 University of Edinburgh, UK.  
Distribution rights by Oxford Molecular Ltd

Message: Protein - protein database search using Smith-Waterman algorithm  
Run on: Tue Jun 20 13:38:51 2000 Master time 4.74 seconds  
408.008 Million cell updates/sec  
Larger output not generated.

File: 05000142524A.p  
Description: (1134) from JSC9142524A.pcf  
Residue Count: 1226  
Sequence: MKVTVAENPGHNRVFIKR.....KFIKRVDSIIAAYQNPASAK 134  
Scoring table: BAW 150  
Baf 11

Search: 145341 seqs, 14437400 residues

Post-processing: Minimum Ratio 0%  
Listing first 45 summaries  
Maximum DB seq length 134

Databases: allseqs  
11342 MB 215B COMB 416 COMB 416 COMB 51backfiles

Statistics: Mean 9.72% Variance 129.31% scale 0.222

Proj. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed.  
and is derived by analysis of the total score distribution.

#### SUMMARY

| Result | Copy | No. | Score | Match length | DB ID | Description | Proj. No. |
|--------|------|-----|-------|--------------|-------|-------------|-----------|
|--------|------|-----|-------|--------------|-------|-------------|-----------|

No matches found.

Search completed: Tue Jun 20 13:38:57 2000  
Job time 1.6 secs.







CC APPLICANT: Bond, Julian F.  
 CC APPLICANT: Garmon, Richard D.  
 CC APPLICANT: Kuo, Mei-Chang  
 CC APPLICANT: Yeung, Siu-mei Hsi  
 CC APPLICANT: Brauer, Andrew  
 CC APPLICANT: Exley, Mark A.  
 CC APPLICANT: Powers, Steven P.  
 CC TITLE OF INVENTION: Allergenic Proteins And Peptides From  
 CC TITLE OF INVENTION: Japanese Cedar Pollen  
 CC NUMBER OF SEQUENCES: 251  
 CC CORRESPONDENCE ADDRESS:  
 CC ADDRESSEE: Immunologic Pharmaceutical Corporation, Inc.  
 CC STREET: 610 Lincoln St  
 CC CITY: Waltham  
 CC STATE: MA  
 CC COUNTRY: USA  
 CC ZIP: 02154  
 CC COMPUTER READABLE FORM:  
 CC MEDIUM TYPE: Floppy disk  
 CC COMPUTER: IBM PC compatible  
 CC OPERATING SYSTEM: PC-DOS/MS-DOS  
 CC SOFTWARE: Patent Release #1.0, Version #1.25  
 CC CURRENT APPLICATION DATA:  
 CC APPLICATION NUMBER: US/98/467,697  
 CC FILING DATE: June 6, 1995  
 CC CLASSIFICATION: 424  
 CC PRIOR APPLICATION DATA:  
 CC APPLICATION NUMBER: 06/150,225  
 CC FILING DATE: December 6, 1994  
 CC ATTORNEY/AGENT INFORMATION:  
 CC NAME: Jane E. Resiliard  
 CC REGISTRATION NUMBER: 38,672  
 CC REFERENCE/DOCKET NUMBER: 025,6 US04 (INT-125024)  
 CC TELECOMMUNICATION INFORMATION:  
 CC TELEPHONE: (617) 227-7400  
 CC TELEFAX: (617) 227-5341  
 CC INFORMATION FOR SEQUENCE NO: 62  
 CC SEQUENCE CHARACTERISTICS:  
 CC LENGTH: 60 amino acids  
 CC TYPE: amino acid  
 CC MOLECULE TYPE: peptide  
 CC FRAGMENT TYPE: integral  
 CC SEQUENCE: 60 AA: 6644 6645 6646 6647 6648 6649 6650 6651 6652 6653 6654 6655 6656 6657 6658 6659 6660 6661 6662 6663 6664 6665 6666 6667 6668 6669 6670 6671 6672 6673 6674 6675 6676 6677 6678 6679 6680 6681 6682 6683 6684 6685 6686 6687 6688 6689 6690 6691 6692 6693 6694 6695 6696 6697 6698 6699 6700 6701 6702 6703 6704 6705 6706 6707 6708 6709 6710 6711 6712 6713 6714 6715 6716 6717 6718 6719 6720 6721 6722 6723 6724 6725 6726 6727 6728 6729 6730 6731 6732 6733 6734 6735 6736 6737 6738 6739 6740 6741 6742 6743 6744 6745 6746 6747 6748 6749 6750 6751 6752 6753 6754 6755 6756 6757 6758 6759 6760 6761 6762 6763 6764 6765 6766 6767 6768 6769 6770 6771 6772 6773 6774 6775 6776 6777 6778 6779 6780 6781 6782 6783 6784 6785 6786 6787 6788 6789 6790 6791 6792 6793 6794 6795 6796 6797 6798 6799 6800 6801 6802 6803 6804 6805 6806 6807 6808 6809 6810 6811 6812 6813 6814 6815 6816 6817 6818 6819 6820 6821 6822 6823 6824 6825 6826 6827 6828 6829 6830 6831 6832 6833 6834 6835 6836 6837 6838 6839 6840 6841 6842 6843 6844 6845 6846 6847 6848 6849 6850 6851 6852 6853 6854 6855 6856 6857 6858 6859 6860 6861 6862 6863 6864 6865 6866 6867 6868 6869 6870 6871 6872 6873 6874 6875 6876 6877 6878 6879 6880 6881 6882 6883 6884 6885 6886 6887 6888 6889 6890 6891 6892 6893 6894 6895 6896 6897 6898 6899 6900 6901 6902 6903 6904 6905 6906 6907 6908 6909 6910 6911 6912 6913 6914 6915 6916 6917 6918 6919 6920 6921 6922 6923 6924 6925 6926 6927 6928 6929 6930 6931 6932 6933 6934 6935 6936 6937 6938 6939 6940 6941 6942 6943 6944 6945 6946 6947 6948 6949 6950 6951 6952 6953 6954 6955 6956 6957 6958 6959 6960 6961 6962 6963 6964 6965 6966 6967 6968 6969 6970 6971 6972 6973 6974 6975 6976 6977 6978 6979 6980 6981 6982 6983 6984 6985 6986 6987 6988 6989 6990 6991 6992 6993 6994 6995 6996 6997 6998 6999 7000 7001 7002 7003 7004 7005 7006 7007 7008 7009 7010 7011 7012 7013 7014 7015 7016 7017 7018 7019 7020 7021 7022 7023 7024 7025 7026 7027 7028 7029 7030 7031 7032 7033 7034 7035 7036 7037 7038 7039 7040 7041 7042 7043 7044 7045 7046 7047 7048 7049 7050 7051 7052 7053 7054 7055 7056 7057 7058 7059 7060 7061 7062 7063 7064 7065 7066 7067 7068 7069 7070 7071 7072 7073 7074 7075 7076 7077 7078 7079 7080 7081 7082 7083 7084 7085 7086 7087 7088 7089 7090 7091 7092 7093 7094 7095 7096 7097 7098 7099 7100 7101 7102 7103 7104 7105 7106 7107 7108 7109 7110 7111 7112 7113 7114 7115 7116 7117 7118 7119 7120 7121 7122 7123 7124 7125 7126 7127 7128 7129 7130 7131 7132 7133 7134 7135 7136 7137 7138 7139 7140 7141 7142 7143 7144 7145 7146 7147 7148 7149 7150 7151 7152 7153 7154 7155 7156 7157 7158 7159 7160 7161 7162 7163 7164 7165 7166 7167 7168 7169 7170 7171 7172 7173 7174 7175 7176 7177 7178 7179 7180 7181 7182 7183 7184 7185 7186 7187 7188 7189 7190 7191 7192 7193 7194 7195 7196 7197 7198 7199 7200 7201 7202 7203 7204 7205 7206 7207 7208 7209 7210 7211 7212 7213 7214 7215 7216 7217 7218 7219 7220 7221 7222 7223 7224 7225 7226 7227 7228 7229 7230 7231 7232 7233 7234 7235 7236 7237 7238 7239 7240 7241 7242 7243 7244 7245 7246 7247 7248 7249 7250 7251 7252 7253 7254 7255 7256 7257 7258 7259 7260 7261 7262 7263 7264 7265 7266 7267 7268 7269 7270 7271 7272 7273 7274 7275 7276 7277 7278 7279 7280 7281 7282 7283 7284 7285 7286 7287 7288 7289 7290 7291 7292 7293 7294 7295 7296 7297 7298 7299 7300 7301 7302 7303 7304 7305 7306 7307 7308 7309 7310 7311 7312 7313 7314 7315 7316 7317 7318 7319 7320 7321 7322 7323 7324 7325 7326 7327 7328 7329 7330 7331 7332 7333 7334 7335 7336 7337 7338 7339 7340 7341 7342 7343 7344 7345 7346 7347 7348 7349 7350 7351 7352 7353 7354 7355 7356 7357 7358 7359 7360 7361 7362 7363 7364 7365 7366 7367 7368 7369 7370 7371 7372 7373 7374 7375 7376 7377 7378 7379 7380 7381 7382 7383 7384 7385 7386 7387 7388 7389 7390 7391 7392 7393 7394 7395 7396 7397 7398 7399 7400 7401 7402 7403 7404 7405 7406 7407 7408 7409 7410 7411 7412 7413 7414 7415 7416 7417 7418 7419 7420 7421 7422 7423 7424 7425 7426 7427 7428 7429 7430 7431 7432 7433 7434 7435 7436 7437 7438 7439 7440 7441 7442 7443 7444 7445 7446 7447 7448 7449 7450 7451 7452 7453 7454 7455 7456 7457 7458 7459 7460 7461 7462 7463 7464 7465 7466 7467 7468 7469 7470 7471 7472 7473 7474 7475 7476 7477 7478 7479 7480 7481 7482 7483 7484 7485 7486 7487 7488 7489 7490 7491 7492 7493 7494 7495 7496 7497 7498 7499 7500 7501 7502 7503 7504 7505 7506 7507 7508 7509 7510 7511 7512 7513 7514 7515 7516 7517 7518 7519 7520 7521 7522 7523 7524 7525 7526 7527 7528 7529 7530 7531 7532 7533 7534 7535 7536 7537 7538 7539 7540 7541 7542 7543 7544 7545 7546 7547 7548 7549 7550 7551 7552 7553 7554 7555 7556 7557 7558 7559 7560 7561 7562 7563 7564 7565 7566 7567 7568 7569 7570 7571 7572 7573 7574 7575 7576 7577 7578 7579 7580 7581 7582 7583 7584 7585 7586 7587 7588 7589 7590 7591 7592 7593 7594 7595 7596 7597 7598 7599 7600 7601 7602 7603 7604 7605 7606 7607 7608 7609 7610 7611 7612 7613 7614 7615 7616 7617 7618 7619 7620 7621 7622 7623 7624 7625 7626 7627 7628 7629 7630 7631 7632 7633 7634 7635 7636 7637 7638 7639 7640 7641 7642 7643 7644 7645 7646 7647 7648 7649 7650 7651 7652 7653 7654 7655 7656 7657 7658 7659 7660 7661 7662 7663 7664 7665 7666 7667 7668 7669 7670 7671 7672 7673 7674 7675 7676 7677 7678 7679 7680 7681 7682 7683 7684 7685 7686 7687 7688 7689 7690 7691 7692 7693 7694 7695 7696 7697 7698 7699 7700 7701 7702 7703 7704 7705 7706 7707 7708 7709 7710 7711 7712 7713 7714 7715 7716 7717 7718 7719 7720 7721 7722 7723 7724 7725 7726 7727 7728 7729 7730 7731 7732 7733 7734 7735 7736 7737 7738 7739 7740 7741 7742 7743 7744 7745 7746 7747 7748 7749 7750 7751 7752 7753 7754 7755 7756 7757 7758 7759 7760 7761 7762 7763 7764 7765 7766 7767 7768 7769 7770 7771 7772 7773 7774 7775 7776 7777 7778 7779 7780 7781 7782 7783 7784 7785 7786 7787 7788 7789 7790 7791 7792 7793 7794 7795 7796 7797 7798 7799 7800 7801 7802 7803 7804 7805 7806 7807 7808 7809 7810 7811 7812 7813 7814 7815 7816 7817 7818 7819 7820 7821 7822 7823 7824 7825 7826 7827 7828 7829 7830 7831 7832 7833 7834 7835 7836 7837 7838 7839 7840 7841 7842 7843 7844 7845 7846 7847 7848 7849 7850 7851 7852 7853 7854 7855 7856 7857 7858 7859 7860 7861 7862 7863 7864 7865 7866 7867 7868 7869 7870 7871 7872 7873 7874 7875 7876 7877 7878 7879 7880 7881 7882 7883 7884 7885 7886 7887 7888 7889 7890 7891 7892 7893 7894 7895 7896 7897 7898 7899 7900 7901 7902 7903 7904 7905 7906 7907 7908 7909 7910 7911 7912 7913 7914 7915 7916 7917 7918 7919 7920 7921 7922 7923 7924 7925 7926 7927 7928 7929 7930 7931 7932 7933 7934 7935 7936 7937 7938 7939 7940 7941 7942 7943 7944 7945 7946 7947 7948 7949 7950 7951 7952 7953 7954 7955 7956 7957 7958 7959 7960 7961 7962 7963 7964 7965 7966 7967 7968 7969 7970 7971 7972 7973 7974 7975 7976 7977 7978 7979 7980 7981 7982 7983 7984 7985 7986 7987 7988 7989 7990 7991 7992 7993 7994 7995 7996 7997 7998 7999 8000 8001 8002 8003 8004 8005 8006 8007 8008 8009 8010 8011 8012 8013 8014 8015 8016 8017 8018 8019 8020 8021 8022 8023 8024 8025 8026 8027 8028 8029 8030 8031 8032 8033 8034 8035 8036 8037 8038 8039 8040 8041 8042 8043 8044 8045 8046 8047 8048 8049 8050 8051 8052 8053 8054 8055 8056 8057 8058 8059 8060 8061 8062 8063 8064 8065 8066 8067 8068 8069 8070 8071 8072 8073 8074 8075 8076 8077 8078 8079 8080 8081 8082 8083 8084 8085 8086 8087 8088 8089 8090 8091 8092 8093 8094 8095 8096 8097 8098 8099 8100 8101 8102 8103 8104 8105 8106 8107 8108 8109 8110 8111 8112 8113 8114 8115 8116 8117 8118 8119 8120 8121 8122 8123 8124 8125 8126 8127 8128 8129 8130 8131 8132 8133 8134 8135 8136 8137 8138 8139 8140 8141 8142 8143 8144 8145 8146 8147 8148 8149 8150 8151 8152 8153 8154 8155 8156 8157 8158 8159 8160 8161 8162 8163 8164 8165 8166 8167 8168 8169 8170 8171 8172 8173 8174 8175 8176 8177 8178 8179 8180 8181 8182 8183 8184 8185 8186 8187 8188 8189 8190 8191 8192 8193 8194 8195 8196 8197 8198 8199 8200 8201 8202 8203 8204 8205 8206 8207 8208 8209 8210 8211 8212 8213 8214 8215 8216 8217 8218 8219 8220 8221 8222 8223 8224 8225 8226 8227 8228 8229 8230 8231 8232 8233 8234 8235 8236 8237 8238 8239 8240 8241 8242 8243 8244 8245 8246 8247 8248 8249 8250 8251 8252 8253 8254 8255 8256 8257 8258 8259 8260 8261 8262 8263 8264 8265 8266 8267 8268 8269 8270 8271 8272 8273 8274 8275 8276 8277 8278 8279 8280 8281 8282 8283 8284 8285 8286 8287 8288 8289 8290 8291 8292 8293 8294 8295 8296 8297 8298 8299 8300 8301 8302 8303 8304 8305 8306 8307 8308 8309 8310 8311 8312 8313 8314 8315 8316 8317 8318 8319 8320 8321 8322 8323 8324 8325 8326 8327 8328 8329 8330 8331 8332 8333 8334 8335 8336 8337 8338 8339 8340 8341 8342 8343 8344 8345 8346 8347 8348 8349 8350 8351 8352 8353 8354 8355 8356 8357 8358 8359 8360 8361 8362 8363 8364 8365 8366 8367 8368 8369 8370 8371 8372 8373 8374 8375 8376 8377 8378 8379 8380 8381 8382 8383 8384 8385 8386 8387 8388 8389 8390 8391 8392 8393 8394 8395 8396 8397 8398 8399 8400 8401 8402 8403 8404 8405 8406 8407 8408 8409 8410 8411 8412 8413 8414 8415 8416 8417 8418 8419 8420 8421 8422 8423 8424 8425 8426 8427 8428 8429 8430 8431 8432 8433 8434 8435 8436 8437 8438 8439 8440 8441 8442 8443 8444 8445 8446 8447 8448 8449 8450 8451 8452 8453 8454 8455 8456 8457 8458 8459 8460 8461 8462 8463 8464 8465 8466 8467 8468 8469 8470 8471 8472 8473 8474 8475 8476 8477 8478 8479 8480 8481 8482 8483 8484 8485 8486 8487 8488 8489 8490 8491 8492 8493 8494 8495 8496 8497 8498 8499 8500 8501 8502 8503 8504 8505 8506 8507 8508 8509 8510 8511 8512 8513 8514 8515 8516 8517 8518 8519 8520 8521 8522 8523 8524 8525 8526 8527 8528 8529 8530 8531 8532 8533 8534 8535 8536 8537 8538 8539 8540 8541 8542 8543 8544 8545 8546 8547 8548 8549 8550 8551 8552 8553 8554 8555 8556 8557 8558 8559 8560 8561 8562 8563 8564 8565 8566 8567 8568 8569 8570 8571 8572 8573 8574 8575 8576 8577 8578 8579 8580 8581 8582 8583 8584 8585 8586 8587 8588 8589 8590 8591 8592 8593 8594 8595 8596 8597 8598 8599 8600 8601 8602 8603 8604 8605 8606 8607 8608 8609 8610 8611 8612 8613 8614 8615 8616 8617 8618 8619 8620 8621 8622 8623 8624 8625 8626 8627 8628 8629 8630 8631 8632 8633 8634 8635 8636 8637 8638 8639 8640 8641 8642 8643 8644 8645 8646 8647 8648 8649 8650 8651 8652 8653 8654 8655 8656 8657 8658 8659 8660 8661 8662 8663 8664 8665 8666 8667 8668 8669 8670 8671 8672 8673 8674 8675 8676 8677 8678 8679 8680 8681 8682 8683 8684 8685 8686 8687 8688 8689 8690 8691 8692 8693 8694 8695 8696 8697 8698 8699 8700 8701 8702 8703 8704 8705 8706 8707 8708 8709 8710 8711 8712 8713 8714 8715 8716 8717 8718 8719 8720 8721 8722 8723 8724 8725 8726 8727 8728 8729 8730 8731 8732 8733 8734 8735 8736 8737 8738 8739 8740 8741 8742 8743 8744 8745 8746 8747 8748 8749 8750 8751 8752 8753 8754 8755 8756 8757 8758 8759 8760 8761 8762 8763 8764 8765 8766 8767 8768 8769 8770 8771 8772 8773 8774 8775 8776 8777 8778 8779 8780 8781 8782 8783 8784 8785 8786 8787 8788 8789 8790 8791 8792 8793 8794 8795 8796 8797 8798 8799 8800 8801 8802 8803 8804 8805 8806 8807 8808 8809 8810 8811 8812 8813 8814 8815 8816 8817 8818 8819 8820 8821 8822 8823 8824 8825 8826 8827 8828 8829 8830 8831 8832 8833 8834 8835 8836 8837 8838 8839 8840 8841 8842 8843 8844 8845 8846 8847 8848 8849 8850 8851 8852 8853 8854 8855 8856 8857 8858 8859 8860 8861 8862 8863 8864 8865 8866 8867 8868 8869 8870 8871 8872 8873 8874 8875 8876 8877 8878 8879 8880 8881 8882 8883 8884 8885 8886 8887 8888 8889 8890 8891 8892 8893 8894 8895 8896 8897 8898 8899 8900 8901 8902 8903 8904 8905 8906 8907 8908 8909 8910 8911 8912 8913 8914 8915 8916 8917 8918 8919 8920 8921 8922 8923 8924 8925 8926 8927 8928 8929 8930 8931 8932 8933 8934 8935 8936 8937 8938 8939 8940 8941 8942 8943 8944 8945 8946 8947 8948 8949 8950 8951 8952 8953 8954 8955 8956 8957 8958 8959 8960 8961 8962 8963 8964 8965 8966

```

CC CC COUNTRY: USA
CC CC ZIP: 02154
CC CC COMPUTER READABLE FORM:
CC CC MEDIUM TYPE: Floppy disk
CC CC COMPUTER: IBM PC compatible
CC CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC CC SOFTWARE: Patent in Release #1.0, Version #1.25
CC CC CURRENT APPLICATION DATA:
CC CC APPLICATION NUMBER: US/08/455,940
CC CC FILING DATE: June 6, 1995
CC CC CLASSIFICATION: 424
CC CC PRIOR APPLICATION DATA:
CC CC APPLICATION NUMBER: 08/350,225
CC CC FILING DATE: December 6, 1994
CC CC APPLICATION NUMBER: 08/225,248
CC CC FILING DATE: April 9, 1994
CC CC APPLICATION NUMBER: 07/939,990
CC CC FILING DATE: September 1, 1992
CC CC APPLICATION NUMBER: PCI/US93/00139
CC CC FILING DATE: January 13, 1993
CC CC ATTORNEY/AGENT INFORMATION:
CC CC NAME: Darlene A. Vanstone
CC CC REGISTRATION NUMBER: 35,725
CC CC REFERENCE/DOCKET NUMBER: 025.6 US (IMI-C28CP2)
CC CC TELECOMMUNICATION INFORMATION:
CC CC TELEPHONE: (617) 465-6000
CC CC TELEFAX: (617) 465-6040
CC CC INFORMATION FOR SEQ ID NO: 62:
CC CC SEQUENCE CHARACTERISTICS:
CC CC LENGTH: 60 amino acids
CC CC TYPE: amino acid
CC CC TOPOLOGY: linear
CC CC MOLECULE TYPE: peptide
CC CC FRAGMENT TYPE: internal
CC CC SEQUENCE 60 AA: 6644 MW: 19464 CN:
CC
CC Query Match 28.8% Score 296: DB 10: Length 60:
CC Best Local Similarity 73.6% Pred. No. 3,14e-18:
CC Matches 39: Conservative 5: Mismatches 7: Indels 2: Gaps 2:
CC
CC 6 RPLWIFSGNNMNIKLMPMYIAGYKTFDGRGAV-YINGGPCVFKRVSNI 57
CC 73 RPLWIFSGNNMNIKLMPMYIAGYKTFDGRRAEVSYVHNGA-KFIRRVGII 124
CC
CC RESULT 10
CC ID US-08-457-023-62 STANDARD: PRI: 50 AA
CC XX xxxxxx
CC
CC Sequence 62: Application US/08457023
CC
CC Sequence 62: Application US/08457023
CC GENERAL INFORMATION:
CC APPLICANT: Griffith, Irwin J.:
CC APPLICANT: Pollock, Joanne:
CC APPLICANT: Bond, Julian F.:
CC APPLICANT: Garman, Richard D.:
CC APPLICANT: Kuo, Mei-Chang:
CC APPLICANT: Yeung, Siu-mei H.:
CC APPLICANT: Brauer, Andrew:
CC APPLICANT: Exley, Mark A.:
CC APPLICANT: Powers, Steven P.:
CC TITLE OF INVENTION: Allergenic Proteins And Peptides From
CC TITLE OF INVENTION: Japanese Cedar Pollen
CC NUMBER OF SEQUENCES: 261
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: ImmuLogic Pharmaceutical Corporation, Inc.
CC STREET: 610 Lincoln St
CC CITY: Waltham
CC STATE: MA
CC COUNTRY: USA
CC ZIP: 02154
CC
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patent in Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/455,940
CC FILING DATE: June 6, 1995
CC CLASSIFICATION: 424
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/350,225
CC FILING DATE: December 6, 1994
CC APPLICATION NUMBER: 08/225,248
CC FILING DATE: April 9, 1994
CC APPLICATION NUMBER: 07/939,990
CC FILING DATE: September 1, 1992
CC APPLICATION NUMBER: PCI/US93/00139
CC FILING DATE: January 13, 1993
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Darlene A. Vanstone
CC REGISTRATION NUMBER: 35,725
CC REFERENCE/DOCKET NUMBER: 025.6 US (IMI-C28CP2)
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (617) 465-6000
CC TELEFAX: (617) 465-6040
CC INFORMATION FOR SEQ ID NO: 62:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 60 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC FRAGMENT TYPE: internal
CC SEQUENCE 60 AA: 6644 MW: 19464 CN:

```

```

CC CC COUNTRY: USA
CC CC ZIP: 02154
CC CC COMPUTER READABLE FORM:
CC CC MEDIUM TYPE: Floppy disk
CC CC COMPUTER: IBM PC compatible
CC CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC CC SOFTWARE: Patent in Release #1.0, Version #1.25
CC CC CURRENT APPLICATION DATA:
CC CC APPLICATION NUMBER: US/08/455,940
CC CC FILING DATE: June 6, 1995
CC CC CLASSIFICATION: 424
CC CC PRIOR APPLICATION DATA:
CC CC APPLICATION NUMBER: 08/350,225
CC CC FILING DATE: December 6, 1994
CC CC ATTORNEY/AGENT INFORMATION:
CC CC NAME: Jane E. Retillard
CC CC REGISTRATION NUMBER: 38,872
CC CC REFERENCE/DOCKET NUMBER: 025.6 USD2 (IMI-C28CP2)
CC CC TELECOMMUNICATION INFORMATION:
CC CC TELEPHONE: (617) 227-7400
CC CC TELEFAX: (617) 227-5941
CC CC INFORMATION FOR SEQ ID NO: 62:
CC CC SEQUENCE CHARACTERISTICS:
CC CC LENGTH: 60 amino acids
CC CC TYPE: amino acid
CC CC TOPOLOGY: linear
CC CC MOLECULE TYPE: peptide
CC CC FRAGMENT TYPE: internal
CC CC SEQUENCE 60 AA: 6644 MW: 19464 CN:
CC
CC Query Match 28.8% Score 296: DB 10: Length 60:
CC Best Local Similarity 73.6% Pred. No. 3,14e-18:
CC Matches 39: Conservative 5: Mismatches 7: Indels 2: Gaps 2:
CC
CC 6 RPLWIFSGNNMNIKLMPMYIAGYKTFDGRGAV-YINGGPCVFKRVSNI 57
CC 73 RPLWIFSGNNMNIKLMPMYIAGYKTFDGRRAEVSYVHNGA-KFIRRVGII 124
CC
CC RESULT 11
CC ID US-08-350-225-187 STANDARD: PRI: 128 AA
CC XX xxxxxx
CC
CC Sequence 187: Application US/08350225
CC GENERAL INFORMATION:
CC APPLICANT: Griffith, Irwin J.:
CC APPLICANT: Pollock, Joanne:
CC APPLICANT: Bond, Julian F.:
CC APPLICANT: Garman, Richard D.:
CC APPLICANT: Kuo, Mei-Chang:
CC APPLICANT: Yeung, Siu-mei H.:
CC APPLICANT: Brauer, Andrew:
CC APPLICANT: Exley, Mark A.:
CC APPLICANT: Powers, Steven P.:
CC TITLE OF INVENTION: Allergenic Proteins And Peptides From
CC TITLE OF INVENTION: Japanese Cedar Pollen
CC NUMBER OF SEQUENCES: 261
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: ImmuLogic Pharmaceutical Corporation, Inc.
CC STREET: 610 Lincoln St
CC CITY: Waltham
CC STATE: MA
CC COUNTRY: USA
CC ZIP: 02154
CC
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS

```







\*\*\*\*\*  
WIREP (TM)  
\*\*\*\*\*

Release 4.1A John F. Collins, BioComputing Research Unit  
Copyright (c) 1993-1998 University of Edinburgh U.K.  
Distribution rights by Oxford Molecular Ltd

MSrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Mon Jun 19 16:01:05 2000 MasPar time 11.24 Seconds  
562,380 Million cell updates/sec

Regular output not generated.

Title: US-09-142-524A-3  
Description: (1-134) from US09142524A.per  
Perfect Score: 1026  
Sequence: 1 MKVTVAFKSGFGRNRVFIKK.....KPIRWVDGIIAAYGNPASK 134

Scoring table: BAX 150  
Gap 12

Searched: 142680 seqs, 47172406 residues

Post-processing: Minimum Match 08  
Listing first 45 summaries  
Maximum DP seq length 134

Database: Firs2  
1:pir1 2:pir2 3:pir3 4:p 4

Statistics Mean 43.199 Variance 89.18 Scale 0.484

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARY

| Result No. | Score | Match | Length | ID     | Description          | Pred. No. |
|------------|-------|-------|--------|--------|----------------------|-----------|
| 1          | 105   | 19    | 2      | D48562 | hypothetical nucleic | 9,82e-02  |
| 2          | 96    | 9     | 1      | F75052 | hypothetical protein | 1,36e-00  |
| 3          | 91    | 8     | 1      | I40656 | hypothetical protein | 5,45e-00  |

Note: Post-processor removed 42 summaries from list due to search parameters chosen.

ALIGNMENTS

RESULT 1  
ENTRY 1  
TITLE 1  
ORGANISM 1  
DATE 1  
ACCESSIONS 1  
REFERENCE 1  
#authors 1  
#journal 1  
#title 1

D48562 #type complete  
hypothetical nucleic acid-binding protein - San Miguel sea  
lion virus (serotype 4)  
#formal\_name San Miguel sea lion virus  
17-Feb-1994 #sequence\_revision 17-Feb-1994 #text\_change  
20-Mar-1998

D48562  
A48562  
Nell, J.D.  
Virus Res. (1992) 24:211-222  
Nucleotide sequence of the capsid protein gene of two  
serotypes of San Miguel sea lion virus: identification of

conserved and non-conserved amino acid sequences about  
calicivirus capsid proteins.

cross-references MJD:9241000  
accession D48562

#status preliminary

#molecule\_type genomic RNA

#residues 1-110 #label NE1

#cross-references DB:M87422; NID:9334866; PIR:G33489

#note sequence extracted from NCBI backbone (NID:933489)  
#length 110 #molecular\_weight 12565 #checksum 27

SUMMARY

Query Match 10/28 Score 105.18 28 Length 110  
Best Local Similarity 31.99 E102. N3 9 82e-02  
Matches 15: Conservative 15; Mismatches 14; Indels 3; Gaps 2

DE 10 PUNSVANVVEGKLDL-ASKGLQKSRALDIERAFNYDR--LAFK 51

QY 17 FKKVSNVILGRRICIFASKNFELKNTIGTGRWKNKRINLQFAK 6

RESULT 2

ENTRY 2

TITLE 2

ORGANISM 2

DATE 2

ACCESSIONS 2

REFERENCE 2

#authors 2

#submission 2

#description 2

#accession 2

#status preliminary

#molecule\_type DNA

#residues 2-111 #label KAM

#cross-references GR:AC248267; GR:AC268866; NID:3545857  
PIONCAPS19151; PIR:e1515213; PIR:3545857

#experimental\_source strain Orsay

GENETICS

#gene

SUMMARY

Query Match 10/28 Score 105.18 28 Length 111

Best Local Similarity 31.99 E102. N3 9 82e-02

Matches 15: Conservative 15; Mismatches 14; Indels 3; Gaps 2

DE 50 KVKLFESTAEVNIIOGAAMAYSMIADIERALIRFNVY-VYLCRGNVWVLSK 109

QY 56 KWLQFALITGFTLMGRRL-WLFSQNMN:KIKMPYIAGYKIFDGRRAEVYVAVNJA 114

DB 109 K 109

QY 115 K 115

RESULT 3

ENTRY 3

TITLE 3

ORGANISM 3

DATE 3

ACCESSIONS 3

REFERENCE 3

#authors 3

#journal 3

#title 3

#cross-references MJD:9435C801

#accession 140656

#status preliminary; translated from GR/EMBL/NCBI



\*\*\*\*\*  
 W E B E R  
 (TM)  
 \*\*\*\*\*

Release 4.1A John F. Collins, BioComputing Research Unit,  
 Copyright (c) 1993-1998 University of Edinburgh, U.K.  
 Distribution rights by Oxford Molecular Ltd

MSRCB\_PP protein - protein database search, using Smith-Waterman algorithm

Run on: Mon Jun 19 16:00:24 2000 Maspar time 7.92 seconds  
 515,241 Million cell updates/sec

Tabular output not generated.

Title: US-09-142-524A-3  
 Description: (-134) from US09-42524A.pef  
 Perfect Score: 126  
 Sequence: 1 MKVTVAENQSPNRRVFIR.....KEIFRVDEGIIAAYQNPASWK 134

Scoring table: B6M 150  
 Gap 11

Searched: 8887 seqs, 3045473 residues

Post-processing: Minimum Match 0%  
 Listing first 45 summaries  
 Maximum LH seq length 134

Database: swiss-prot38  
 trswissprot

Statistics: Mean 44.072; Variance 79.368; scale 0.555

Pred. No. is the number of results predicted by chance to have a  
 score greater than or equal to the score of the result being printed,  
 and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score | Match | Length | DB ID | Description                       | Pred. No. |
|------------|-------|-------|--------|-------|-----------------------------------|-----------|
| 1          | 105   | 100   | 110    | 1     | Y12K_SMSV4 POSSIBLE 12 KD NUCLEIC | 1.59e-02  |

Note: Post-processor removed 44 summaries from list due to search parameters chosen.

#### ALIGNMENTS

RESULT 1  
 ID Y12K\_SMSV4 STANDARD: PRT: 110 AA.  
 AC P36289;  
 DT 01-JUN-1994 (Rel. 29, Created)  
 DI 01-JUN-1994 (Rel. 29, Last sequence update)  
 DI 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE POSSIBLE 12 KD NUCLEIC ACID-BINDING PROTEIN.  
 OS San Miguel sea lion virus (serotype 4) (SMSV 4).  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage: Caliciviridae;  
 CC Calicivirus.  
 RN 111  
 RF SEQUENCE FROM N.A.  
 RX MEDLINE: 92410750.  
 RA Nelson J.D.;  
 RT "Nucleotide sequence of the capsid protein gene of two serotypes of

RT San Miguel sea lion virus: identification of conserved and non-  
 conserved amino acid sequences among calicivirus capsid proteins  
 RL Virus Res. 24:211-222(1992).  
 CC -!- SIMILARITY TO FELINE CALICIVIRUS 12 KD PROTEIN.  
 CC  
 CC This SW-SS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Data Bank  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/submit/submit.html>)  
 CC or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk).  
 CC  
 CC EMBL: M67482; AAA16021.1;  
 CC PIR: D48562; S49362.  
 CC  
 CC SEQUENCE 110 AA: 12566 MW. 14255559482741EE 09464  
 Query Match 10.2%; Score 105; DB 1; Length 110  
 Best Local Similarity 31.9%; Pwd No. 1.59e-02  
 Matches 15; Conservative 15; Mismatches 14; Indels 4; Gaps 2;  
 Db 10 FANSVANAVVEGKKDL-ASKGLQKNSRALDTERAFNYIH--IAFEK 53  
 QY 17 FKKVSNVILHGRRIDIFASKNPHCKNTIGTGRWKNKPNWLQFAK 64  
 Search completed: Mon Jun 19 16:00:34 2000  
 Job time : 10 secs.

\*\*\*\*\*  
 WAREHOUSE  
 \*\*\*\*\*

Released via John P. Collins, BioComputing Research Unit,  
 Copyright (c) 1993-1998 University of Edinburgh, U.K.  
 Distribution Rights by Oxford Molecular Ltd

Miscellaneous protein database search, using Smith-Waterman algorithm  
 Run on: Mon Jun 19 16:00:52 2000; Maspar time 18.49 Seconds  
 502.347 Million cells updates/sec  
 Tabular output not generated.

Title: US-09-142-524A-3  
 Description: (1114); from US09142524A.pep  
 Perfect Score: 1.26  
 Sequence: MKVIVAFNPGPRKRVFIR.....KPIRVVDGIIAAYQNPASWK 134

Scoring matrix: BAW 150  
 Gap 12

Searched: 125876 seqs, 69334122 residues

Fast-processing: Minimum Match 0%  
 Listing first 45 summaries  
 Maximum DB seq length 134

Database: 41000012  
 1. sp\_archaea 2. sp\_bacter 3. sp\_fungi 4. sp\_human  
 5. sp\_invertebrate 6. sp\_mammal 7. sp\_phc 8. sp\_orquehelie  
 9. sp\_plant 10. sp\_plant 11. sp\_rodent 12. sp\_unclassified  
 13. sp\_vertebrate 14. sp\_virus

Statistics: Mean 42.980; Variance 75.878; scale 0.566

Pred No. is the number of results predicted by chance to have a  
 score greater than or equal to the score of the result being printed,  
 and is derived by analysis of the total score distribution.

# SUMMARIES

| Result No. | Query Match | DB ID | Description         | Pred. No. |
|------------|-------------|-------|---------------------|-----------|
| 1          | 95          | 10814 | CAPSID              | 5.52e-01  |
| 2          | 90          | 551   | PROTEIN TRANSLOCASE | 2.76e-00  |

Note: Post-processor removed 43 summaries from list due to search parameters chosen.

# ALIGNMENTS

RESULT 1  
 ID C09783 PRELIMINARY: PRT: 108 AA.  
 AC C09783  
 DI 01-JUL-1997 (TrEMBLrel. 04, Created)  
 DI 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)  
 DI 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)  
 DE CAPSID.  
 OS Saccharomyces cerevisiae virus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Caliciviridae;  
 CC Calicivirus.  
 RN [1]

RP SEQUENCE FROM N.A.  
 RC STRAIN=SMSV-17;  
 RX MEDLINE: 98187296;  
 RA WATSON D.O., BESKE T., GINSBURG M.B., POST S.E., ZHENG W.X., ZHANG X.M.,  
 RA JIANG X., GOLDING B., SMITH A.W.;  
 RT "Partial characterization of the genome of nine calici-  
 RT caliciviruses".  
 RL Arch. Virol. 141:2443-2456(1996).  
 CR EMBL: U52005; AAC57041.1;  
 SQ SEQUENCE 108 AA: 12354 MW: 4763845 SPC92

Query Match 9.4% Score 90; DB 14; Length 108  
 Best Local Similarity 34.0% Prof. No. 5.52e-01  
 Matches 16; Conservative 16; Mismatches 15; Indels 2;

DB 10 FINGERVANALISCKKIDLASDKNKALITKKNYARSLAFKFA  
 QY 17 EIKRVSNVIMHGRIDIPASDQKANTGUTHRKKNINLOFAKFA

RESULT 2  
 ID C29714 PRELIMINARY: PRT: 55 AA.  
 AC C29714

DI 01-JAN-1996 (TrEMBLrel. 05, Created)  
 DI 01-JAN-1996 (TrEMBLrel. 05, Last sequence update)  
 DI 01-AUG-1998 (TrEMBLrel. 07, Last annotation update)  
 DE PROTEIN TRANSLOCASE, SUBUNIT SEC61 GAMMA (SET1).  
 GN AF0536.

OS Archaeoglobus fulgidus.  
 OC Archaea; Euryarchaeota; Archaeoglobales; Archaeoglobaceae;  
 CC Archaeoglobus.  
 RN [1]

RP SEQUENCE FROM N.A.  
 RC STRAIN=VC-16 / DSM 4304 / AF0536;  
 RX MEDLINE: 98049343;  
 RA KLECK H.P., CLAYTON R.A., JUMP J.P., WHITE O., NELSON K.B.,  
 RA KETCHUM K.A., JOHNSON R.A., OWEN M., HICKY E.K., PETERS N.C.,  
 RA RICHARDSON D.L., KERVAGE A.R., GRAHAM D.E., KISTLER N.C.,  
 RA FLEUSCHMANN R.D., QUACKENBUSH J., LEE N.H., SUTIN N.B., GILES S.  
 RA KIRKNESS E.F., DOUGHERTY D.A., WKENNEY K., ADAMS M.D., LEE S.  
 RA PETERSON S., REICH C.J., MONTEILLO K., BAKER J.H., GIERER A., ZHANG L.  
 RA OVERBEERK R., GOCAYNE J.D., WEIDMAN J.F., McDONALD L., UTTERBACK T.  
 RA COTTON M.D., SPRIGGS T., ALLACH P., KAINE B.P., SYKES S.M.,  
 RA SADOWSKA P., D'ANDREA K.P., KOKMAN C., FULLI C., GARLAND S.A.,  
 RA WASTENHIM, OLSEN G., F. SER C.M., SMITH R., WISE C.  
 RA VENTER J.C.

RC The complete genome sequence of the hyperthermophilic, obligate  
 RC reducing archaeon Archaeoglobus fulgidus.  
 PL Nature 392:694-700(1999).

CR EMBL: AAC01067; AAB47044.1;  
 CR TIGR: AF0536;  
 KW Hypothetical protein.  
 SQ SEQUENCE 55 AA: 7547 MW: 93.887EA CRC92;

Query Match 8.8% Score 90; DB 14; Length 66;  
 Best Local Similarity 36.7% Pred. No. 2.76e-00;  
 Matches 11; Conservative 7; Mismatches 11; Indels 1; Gaps 1;

DB 24 MAKKPDWEEFSNTAKVALAVMEIVGVGF 52

QY 70 KMRRLPWTIFSONMNIKMKMPYAGYKF 99

Search completed: Mon Jun 19 16:01:18 2000  
 Job time : 26 secs.